Social Science Library

APR 15 1954

Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
AIR FORCE PERSONNEL AND TRAINING RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Cletus J. Burke, *Indiana University*
Robert M. Gagné, *AF Personnel and Training Research Center*
W. R. Garner, *The Johns Hopkins University*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *AF Personnel and Training Research Center*
Arthur L. Irion, *Tulane University*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Edwin B. Newman, *Harvard University*
Kenneth W. Spence, *State University of Iowa*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

CONTENTS

Acquisition of S-R Connections: A Test of Hull's and Guthrie's Theories: V. W. VOEKS .. 137
Visual Distance Discrimination in the Rat: A. GREENHUT 148
Transfer in Verbal Materials with Dissimilar Stimuli and Response Similarity Varied: R. K. YOUNG AND B. J. UNDERWOOD 153
The Role of Overt Errors in Serial Rote Learning: H. SCHEIBLE AND B. J. UNDERWOOD ... 160
Factors Influencing Incidental Learning: G. H. BROWN 163
Incidental Learning under Two Incentive Conditions: H. P. BAHRICK 170
Learning Response Compounds Having Two Critical Components: A. J. NORTH AND J. HARRINGTON, JR. 173
Studies of the Effect of Change of Drive: III. Amounts of Switching Produced by Shifting Drive from Thirst to Hunger and from Hunger to Thirst: H. H. KENDLER, A. D. KARASIK, AND A. M. SCHRIER 179
Response Termination of the Cue Stimulus in Classical and Instrumental Conditioning: D. D. WICKENS AND C. E. PLATT 183
The Nature of Errors in Experimental Lie Detection: D. VAN BUSKIRK AND F. L. MARCUSE 187
The Role of Auditory Localization in Attention and Memory Span: D. E. BROADBENT 191
Affectivity as a Factor in the Apparent Size of Pictured Food Objects: H. L. BEAMS 197
Accuracy of Response as a Function of Target Width: E. A. BILODEAU 201
Accuracy of Recognition with Alternatives before and after the Stimulus: D. H. LAWRENCE AND G. R. COLES 208
The Role of Proprioceptive Feedback in Positioning Responses: B. WEISS 215

American Psychological Association

Vol. 47 No. 3 March 1954

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc. The subscription rate per volume is $7.50, or $15.00 annually. Single copies are $1.50. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1954 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Air Force Personnel and Training Research Center, Lackland Air Force Base, San Antonio, Texas.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 (4, Suppl.), 389–449. In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879.

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Copyright, 1954, by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 47, NO. 3 MARCH, 1954

ACQUISITION OF S–R CONNECTIONS: A TEST OF HULL'S AND GUTHRIE'S THEORIES

VIRGINIA W. VOEKS

Yale University[1]

Is an S–R connection gradually strengthened by repeated reinforcements? Or is an S–R connection established suddenly, in an all-or-none fashion?

Hull has given us the most complete statement of the view that the connections are gradually strengthened. His most recent theory states (**10**, p. 5): "Whenever an effector activity is closely associated with a stimulus afferent impulse or trace (s) and the conjunction is closely associated with the rapid diminution in the motivation stimulus (S_D or S_G), there will result an increment to a tendency for that stimulus to evoke that response" (Postulate 3). "If the reinforcements follow each other at evenly distributed intervals, everything else constant, the resulting habit will increase in strength as a positive growth function of the number of trials" (Postulate 4).

Guthrie has offered the alternative suggestion that an S–R connection is fully established in one pairing of the stimuli and the new response (cf. **5**, p. 419 f.; **6**, p. 98).

The experiment here reported attempts to test the validity of these alternative proposals.

Conditions Required by Hull's Theory

Although S–R connections cannot be observed directly, their nature can be inferred under certain circumstances implicit in the two theories. In Hull's theory, we can infer the nature of S–R if we ascertain the value of ${}_sE_R$ and the corresponding ${}_sH_R$. To determine these values and how they vary with N, the following conditions are necessitated by Hull's 1952 postulates. They were necessitated also by his 1943 theory, though for slightly different reasons.

Requirement 1: Stimulus-variability from trial to trial must be negligible.— Meeting this condition stabilizes V, desirable in terms of Postulate 8.[2] More important, it minimizes variations in afferent interaction, so ${}_sE_R$ has a relatively determinable magnitude.

Hull assumes *all* afferent impulses (s's) present at any time interact and partially change each other (into s's), thus changing ${}_sE_R$ (Postulate 11). Unfortunately,

[1] This paper is based upon parts of a doctoral dissertation presented to the faculty of Yale University. The author is now with San Diego State College.

[2] To reduce printing costs, Hull's postulates are referred to by number only. Complete statements are available elsewhere (**10**, pp. 5–14).

he does not specify the empirical conditions under which greater or lesser changes from afferent interaction occur. Suppose the stimuli on Trials 4, 5, and 6 differ somewhat. Will the stimuli on Trial 5 interact to give $\bar{s}$'s more like the $\bar{s}$'s of Trial 4 than those resulting from interaction of stimuli present on Trial 6? Or less? Or the same? At present we cannot answer those questions without circular assumptions. Consequently, when different stimuli are present on various trials, we cannot predict either the amount or direction in which the resultant ${}_{S}E_{R}$'s will differ from trial to trial.

Requirement 2: Stimuli present on trials (S_t) should be absent between trials.—Meeting this condition avoids confounding reinforced and nonreinforced S_t-R_+ sequences. It also minimizes I_{R+} and ${}_{S_t}I_{R+}$, which should be negligible since they reduce ${}_{S_t}E_{R+}$ by unknown amounts (Postulate 9A and Corollary 9).

Although the theory suggests conditions under which the various inhibitory potentials increase, it does not enable us to predict *how much* they increase under various conditions, but only *that* they increase. Consequently, we cannot predict the nature of the response-evocation curve when appreciable amounts of the inhibitory potentials are present.

Any results can be postdicted when Requirements 1 and 2 are not fulfilled, by assigning (legitimately) different values for various trials to the inhibitory potentials and to the effects of afferent interaction. Postdictions under such circumstances afford no check on the theory. This means that to test Postulates 3, 4, or 8, as the theory currently stands, the various inhibitory potentials and the changes in afferent interaction must be negligible.

Requirement 3: Ample intertrial time must be allowed for dissipation of I_R.—This follows from the above considerations and from Postulate 9B.

Requirement 4: The response must involve minimal work.—This will further reduce ${}_{S_t}I_{R+}$, since ${}_{S_t}I_{R+}$ is a function in part of the amount of I_{R+} present with the on-trial stimuli (Corollary 9) and the amount of I_{R+} present is hypothesized as being an increasing function of the amount of work involved in making R_+ (Corollary 10).

Requirement 5: Drive and incentive must be as nearly constant as possible.—Unmeasured variations in these factors would cause unpredictable fluctuations in ${}_{S}E_{R}$ (Postulate 8 and Corollary 12), and thus preclude determining the nature of ${}_{S_t}H_{R+}$.

Requirement 6: The original level of ${}_{S_t}E_{R_+}$ should be zero.—This is desirable since otherwise the original level of ${}_{S_t}E_{R+}$, which partially determines the shape of the learning curves, cannot be established.

Conditions Required by Guthrie's Theory

Nearly the same conditions necessitated by Hull's theory are needed to test Guthrie's theorem predicting jumpwise learning curves for stable conditions (Theorem 6, **16**, p. 358). A stable situation was defined as "a set of stimulus-patterns . . . having the following characteristics—(*a*) all stimuli present on Trial X are present on all succeeding trials, (*b*) none of the stimuli present on Trial X is present between trials, (*c*) none absent from Trial X is added on succeeding trials" (**16**, p. 357). Trial X is the first trial on which R_+ occurs; R_+ is the response E wishes to cue to the experimental stimulus-patterns.

Therefore, to test Theorem 6, Requirements 1, 2, and either 3 or 4 must be met. Either the response must involve little work or the time between trials must permit dissipation of fatigue products, since (by Guthrie's theory) the accumulation of fatigue products creates new, varying stimulus-patterns.

METHOD

A conditioning procedure was used, R_+ being a partial or complete blink or wink; US, an air-puff by the fall of 100–130 mm. of mercury, magnetically released; one of the main CS's, a damped buzz. The buzz began about .45 sec.

before the puff and ended about .09 sec. after onset of the puff.

Subjects and Apparatus

All *S*s were male undergraduates of Yale University. None had studied conditioning. In the main experiment, 32 *S*s participated, and 25 others in preliminary tests. The apparatus is substantially the same as that previously described and photographed (13). A continuously recording camera designed by W. R. Miles was used to photograph the eyelid responses (magnified 3.5 times) and the onset and offset of all stimuli.

Conditions

Requirement 1.—To maintain more stable and unique on-trial stimuli, the following procedures were introduced. Throughout each trial, *S* depressed a telegraph key with each hand. The keys were depressed "as soon as possible" after a signal, "with a firm smooth motion," held down "with constant pressure" until a buzz terminated, and then released. Each *S* practiced making the response in the particular way sketched. Depression of both keys activated the buzz; contact was broken if either key was released slightly, thus affording a check on the relative constancy of pressure. A stiff spring was used in the keys, so the response involved muscles of the arms and shoulders as well as of the hands.

At the start of each trial, *S* breathed in rhythm to *E*'s saying "Ready, inhale, exhale, inhale, hold." The words were spoken slowly, with about 2 sec. between the first four, 1½ to 3 sec. between the last two. On the last "inhale," *S* inhaled "as deeply as possible," and in response to "hold," held his breath and depressed the keys.

Besides stabilizing the key-response, this routine should have increased the similarity of the on-trial stimuli arising from the diaphragm, intercostals, and other muscles of the chest, shoulders, and upper arms.

To minimize autonomic changes and concomitant changes in internal stimuli, *S* must have either a constant excitement or a constant calmness throughout the experiment. The latter is more feasible. To reduce the apprehension with which many *S*s start (but usually do not finish) their first experimental session, *E* chatted with *S*, answered his questions, and appeared, at least, to allay his many fears about possible shocks, snakes, and other noxious stimuli. Instructions were given slowly, in a conversational manner. The *S* practiced all parts of his routine and was assured he was doing fine. The experiment began only when *S* appeared comfortable and confident he could do what was expected of him. To mitigate increasing "nervousness" about the need for finishing the session, *S* was asked to allow an hour before he signed up; on arrival, he was told that although it might seem longer, we would stop after no more than 45 min. No *S* was run during his examination week nor immediately before any class test.

Changes from increasing fatigue and also in proprioception were reduced by keeping the session as short as was consistent with other considerations, having the intertrial interval long enough to permit *S* to relax between trials and yet short enough to avoid impatience (as evinced by *S*'s reports and almost no signs of restlessness), adjusting *S*'s chair so he sat in his normal position with his back supported, and arranging the keys so each *S* could rest his arms comfortably on the table between trials.

If *S* speculates about the nature of the experiment, forming first one hypothesis and then another, and adopting first one set and then another toward blinking, this unstabilizes a segment of the on-trial stimulus-patterns. Therefore, *S* was told that this was a study of "individual differences in reaction time." No *S* questioned the stated purpose; most *S*s expressed interest in the results of their "reaction-time test."

A soundproof room was used; all necessary movements by *E* were highly routinized; except for these, *E* was nearly soundless; *S*'s head was surrounded by a box, sharply limiting his visual field; *S* was requested not to talk during the experiment (unless fatigued or something seemed to be going wrong).

Requirement 2.—The following sets of stimuli were present on trials but absent between trials: (*a*) Most of the proprioceptive and tactual stimuli from the key-response; (*b*) stimuli from inhaling deeply and holding one's breath; (*c*) various visual stimuli (*S* closed his eyes between trials); (*d*) the buzz.

Requirement 3.—Intertrial intervals were 45, 50, 55 . . . 75 sec., arranged in randomized order. Though the time needed for I_R to dissipate is not known, two lines of evidence suggest this was ample: (*a*) Probably few fatigue products were present immediately after the response; negligible amounts should remain after blood circulates past the effectors for about a minute. (*b*) Calvin (3) showed that with 20 sec. between trials, probability of the eyelid response was 54% and with 3.33 sec., 20% after equal conditioning trials; these data suggest the I_R dissipation curve is leveling off at 20 sec.

Requirement 4.—The eyelid reaction entailed minimal work.

Requirement 5.—Pressure of the puff was adjusted to each *S*'s sensitivity and elicited practically a full wink on each trial. The puff was directed to the middle of *S*'s eyeball, since puffs to various parts of the eyeball are probably more alike than ones to the eyeball and ones to the cornea, and it is easier to hit the eyeball consistently than to hit the cornea consistently. The *S*'s chin was supported on a chin rest; his head was strapped snugly against a forehead rest; and, during the trials, he looked at a cross, thus further stabilizing the US (and possibly *D* and *K*).

Requirement 6.—To each preliminary *S*, all CS's were presented 30 times without the puff. Photographs also were made of five preliminary tests for each *S* in the main experiment. None of these *S*s showed any trace of a wink to any of the CS's.[3]

Number of Trials

The mean number of trials was 17.1, a range of 9 to 25. The number was kept small to avoid increasing fatigue or restlessness, but was sufficient to be certain learning had occurred. Each *S* made at least five R_+'s (mean, 12.1) and at least 32% of his total trials contained R_+ (mean, 72%).[4]

Criteria for R_+

After studying preliminary records of voluntary blinks and other incidental movements occurring when the cross was fixated for 5 sec., the following criteria were selected to differentiate between CR's and voluntary or incidental responses. An eyelid response was counted as R_+ only when it involved a closing-opening sequence, an amplitude of at least 1½ mm., occurred after onset of the buzz but before the puff, and had a record line of rapid ascent and peakedness. The R_+'s with relatively short and long latency were compared with respect to shape, amplitude, frequency, and changes in amplitude and frequency as a function of number of preceding trials. The two groups of responses were substantially the same: short-latency responses behaved a bit more like what one expects of CR's than did the relatively longer latency responses, but with no statistically significant differences (15). Therefore, all R_+'s are considered together in this study.[5]

Other *E*s have included responses occurring after onset of the puff (4, 11, 12), ones of less than 1½ mm. amplitude (4, 7, 14), flat humps (7), gradual tetanic lowering without raising and gradual tetanic raising of the lid (2). Though these may well have been CR's too, incidental responses (from head movements, shifts in gaze, etc.) sometimes give highly similar records. Therefore, they were not counted as R_+ in this study. Since our criteria are more definite and stringent than what appear to be usual, it is unlikely any greater stability of R_+ here is due to "flexibility" in what was counted.

Results and Discussion

Tests of Hull's Theory

Amplitude of response.—According to Hull's theory, ${}_sE_R$ progressively increases with the number of S–R reinforcements (Postulates 3, 4, and 8). Therefore, if amplitude of eyelid responses is similar to that of the PGR or is positively correlated to any degree with ${}_sE_R$,[6] amplitude of these

[3] The various CS's were tested on *S*s run before the 57 *S*s, and were made so weak it seemed unlikely they would evoke any eyelid responses. However, two people (originally intended as *S*s) blinked within a half-second after a CS, and so were dropped.

[4] For an extinction experiment, both the number of R_+'s per *S* and the percentage of trials containing R_+ had to be equated for members of this group and for members of four other groups who learned under different conditions. This precluded having the same number of trials for all *S*s.

[5] When using procedures similar to the present one, other *E*s also have included as CR's responses which appeared within a few milliseconds after onset of the part of the CS sequence corresponding to my buzz, or even a few milliseconds before its onset (cf. 1).

[6] Literally and precisely construed, no amplitudes (except those of the PGR) can be interpreted within the framework of Hull's theory. Postulate 15 of the 1943 version anchors ${}_sE_R$ to amplitude only of responses mediated by the autonomic nervous system, while Postulate 15 of the 1952 version limits the anchorage to amplitude of the Tarchanoff galvanic skin reaction. Strictly speaking, the present amplitude data are irrelevant to Hull's theory. However, the data are relevant to Hull's theory as used by many investigators. Even Hull himself used mean amplitude of the lid reaction as an indicator of ${}_s\bar{E}_R$ (8, p. 291, p. 306) and in discussions made it parallel to amplitude of PGR (8, p. 291f). Similarly Hull used the amplitude of line-drawing responses as a reflector of ${}_sE_R$ (8, p. 306 f.), and presented curves for the amplitude of bar-pressing responses "to illustrate the oscillation of reaction potential (${}_sE_R$)" (8, p. 305).

responses should increase with N (Postulate 15)—providing ${}_{S}O_{R}$, I_{R}, and ${}_{S}I_{R}$ increase with N no more rapidly than does ${}_{S}E_{R}$, the stimuli present on successive trials are substantially the same, and D, V, and K remain substantially constant. It will be recalled that special effort was made to meet these assumptions. (We have no control over ${}_{S}O_{R}$, of course.)

To test this hypothesis, we calculated the mean amplitude of each S's responses during the various quarters of his trials beginning with the first trial he made R_{+}.[7] The mean for each quarter was compared with every other quarter, making six comparisons for each S. In no case is the percentage of Ss whose data accord with the theory significantly more than 50% by usual statistical standards (see Table 1).

TABLE 1

RELATIVE AMPLITUDE OF RESPONSES IN VARIOUS QUARTERS OF EACH S'S TRIAL BEGINNING WITH HIS FIRST R_{+}

In Accord with Theory		At Variance with Theory		p**
Quarters Compared*	Per Cent of Ss	Quarters Compared	Per Cent of Ss	Sig. of Diff. from 50%
1st<2nd	65.62%	1st>2nd	34.38%	.04
1st<3rd	56.25	1st>3rd	43.75	.24
1st<4th	65.62	1st>4th	34.38	.04
2nd<3rd	59.37	2nd>3rd	40.63	.15
2nd<4th	68.75	2nd>4th	31.25	.02
3rd<4th	62.50	3rd>4th	37.50	.08

* No two quarters for any S had the same mean amplitude.
** One-tailed test of significance.

A second analysis is summarized in Table 2. For 6 of 32 Ss, amplitude behaved as would be expected, the mean of each quarter of trials becoming progressively larger. For no S was the curve completely reversed. By "chance," 1.33 Ss would fall in each of those categories. Although again in the right direction, throughout the distribution the deviations from "chance" are small and not statistically significant.

For 25 of the 32 Ss, the first R_{+} was smaller than the last R_{+}. This is significantly more than half of the Ss ($p<.01$).

Probability of response.—In the last two revisions of his theory, Hull did not formally anchor ${}_{S}E_{R}$ or ${}_{S}H_{R}$ to P_{R}. Presumably he was working on a new equation for this, but had not arrived at a satisfactory one. Clearly he intended to keep the concept, since his last book abounds with theorems, analyses, and theoretical graphs using P_{R+} as an indicator of ${}_{S}E_{R}$ and ${}_{S}\bar{E}_{R}$.

From Hull's discussions (**9, 10**) and

[7] Hull postulates that ${}_{S}H_{R}$ gradually increases with N even while the correlated ${}_{S}E_{R}$ minus ${}_{S}O_{R}$ and I_{R} is less than ${}_{S}L_{R}$. That is, ${}_{S}H_{R+}$ is said to gradually increase with N even before R_{+} once has been made in the presence of those Ss. Maybe it does. However, none of the observable correlates of ${}_{S}H_{R}$ will show any change with N before R_{+} has once appeared. Obviously amplitude, latency, and probability of the never-yet-made R_{+} cannot reflect the presence of the postulated increases in ${}_{S}H_{R}$. Equally obviously they cannot reflect the absence of the postulated increases in ${}_{S}H_{R}$ until the R once has been made. It would be nice to avoid confounding those data which can have no bearing on Hull's theory with the data which might have some bearing. Therefore, in this study, the trials *before* the first R_{+}'s were grouped together, and the trials from the first R_{+} on were considered separately.

For each S, these trials were divided into quarters. Precisely how this was done can be illustrated best by an example. Suppose S made no R_{+} for three trials and then had 21 more trials. Each quarter after and including his first R_{+} contains 5¼ trials. The total amplitude of his first quarter is the sum of amplitudes for Trials 4 to 8 inclusive plus ¼ the amplitude of R_{+} on Trial 9. Similarly the total amplitude for the second quarter is the sum of ¾ the amplitude of R_{+} on Trial 9, plus those from Trials 10 through 13, plus ¼ the amplitude of R_{+} on Trial 14. The total amplitudes for the third and fourth quarters are computed in the same way. Dividing each total by 5¼ gives the mean amplitude per quarter for this S. Mean amplitudes are calculated from those of the largest R_{+} on each trial.

TABLE 2

RELATIONSHIP BETWEEN AMPLITUDE AND N FOR INDIVIDUALS

Measure	Number of Times Mean Amplitude of an Earlier Quarter Was Smaller Than That of Some Later Quarter*						
	0	1	2	3	4	5	6
Number of *Ss*	0	4	4	6	5	7	6
Number expected by "chance"	1.33	4.00	6.67	8.00	6.67	4.00	1.33

$\chi_r^2 = 3.82$ (combining categories 0–2 and 4–6); 2 *df*; $p = .16$

* No two quarters had the same mean amplitude.

Postulates 3 and 4, we would expect that under the present conditions S would make R_+ with a frequency gradually increasing with N.

This expectation is not confirmed. Each of 16 *Ss* invariably made R_+ after once doing so. The data for the other 16 *Ss* also offer little evidence that ${}_sE_R$ gradually increases with N. For these *Ss* too, the on-trial stimuli almost invariably evoked R_+ after it appeared once. In other words, most *Ss* had horizontal "learning curves" with a dip perhaps in one place. Table 3 presents a comparison of each quarter with every other quarter of an S's trials *following* his first R_+.[8] For no S was P_{R+} progressively larger from quarter to quarter. Only six *Ss* had data resembling a progressive increase. The nature of these curves can be seen from Table 4. Each has a shape quite different from the others.

Tests of Guthrie's Theory

Incidence of jumpwise curves.—In stable situations, an S who has once made a new response should make that response on every subsequent trial, regardless of the past frequency of other responses (Theorem 6). A stable situation is so defined that stimuli can be dropped from trial to trial, but new stimuli cannot be added nor can on-trial stimuli be present between trials, without unstabilizing the stimulus-series.

In progressively more unstable situations, the curves will deviate from a jumpwise character to a correspondingly greater extent. The complete rationale of this proposition is given in Theorems 7 and 8 (**16**, p. 359 f.).

In this experiment, stimulus-stability was high, but not complete. Hence we should predict from the theory that not all *Ss* will have jumpwise curves. However, the majority of individual curves should closely approximate a jumpwise character. If they do not, it is probable the theory is invalid.

Of the 32 *Ss*, 16 had perfect jump-

TABLE 3

RELATIVE P_{R+} IN VARIOUS QUARTERS OF EACH S'S TRIALS AFTER HIS FIRST R_+

In Accord with Theory		At Variance with Theory	
Quarters Compared	Per Cent of *Ss*	Quarters Compared	Per Cent of *Ss*
1st < 2nd	15.62%	1st ≧ 2nd	84.38%
1st < 3rd	18.75	1st ≧ 3rd	81.25
1st < 4th	21.87	1st ≧ 4th	78.13
2nd < 3rd	15.62	2nd ≧ 3rd	84.38
2nd < 4th	21.88	2nd ≧ 4th	78.12
3rd < 4th	18.75	3rd ≧ 4th	81.25

[8] Including the trial on which R_+ first occurred would have worked against Hull's theory, since then the first quarter would have had to contain at least one R_+.

wise curves. Most of the others had curves closely approximating that: 10 *Ss* had only one lapse; 4 had two lapses; 2 had three or four.

Predictions of individual responses for each trial.—Whatever response an *S* made on the immediately preceding trial was predicted for the next trial. This particular way of using the principle of postremity assumes that the stimuli present at the time for which the response is being predicted were present on some former occasion when the response now predicted had been evoked and were not subsequently present during an incompatible response.

There is a far more precise way to use the principle of postremity and Guthrie's other postulates: Note as many as possible of the stimuli now present and what response was underway when each of those stimulus-patterns was last present. Then predict as the response whatever response the majority of the current stimuli last accompanied, whether or not that was the response made on the immediately preceding trial. This method of prediction, through a more refined use of the theory, would tend to vitiate this experiment, for it quite naturally would raise a question concerning subjectivity. We therefore followed the rather crude, but objective, method sketched above.

TABLE 4

CHANGES IN P_{R_4} FROM FIRST TO LAST QUARTER FOR *Ss* MOST NEARLY FITTING HULLIAN EXPECTATIONS

Patterns Resembling a Progressive Increase of P_{R_4} with N	Number of *Ss* with Each Pattern
1st = 2nd = 3rd < 4th	1
1st = 2nd < 3rd = 4th	1
1st < 2nd = 3rd = 4th	2
1st < 2nd < 3rd = 4th	2
1st < 2nd < 3rd < 4th	0
1st ≦ 2nd ≦ 3rd ≦ 4th	6 Total

TABLE 5

ACCURACY OF PREDICTIONS FROM GUTHRIE'S THEORY FOR INDIVIDUAL RESPONSES

Predicted Correctly	Number of *Ss*	p*
90–96%	15	<.006 for each *S*
80–89	9	<.01 for 8 *Ss*, .03 for 1 *S*
70–79	5	<.03 for 3 *Ss*, <.05 for 1 *S*, .07 for 1 *S*
60–69	2	.15 for 1 *S*, .21 for 1 *S*
50–59	1	.50 for 1 *S*

* One-tailed test of significance, using the binomial expansion.

The percentage of individual responses predicted correctly from Guthrie's theory and the statistical significance for each *S* is summarized in Table 5. For almost all *Ss*, the predictions were highly accurate: over 90% of the predictions were correct for 15 of the 32 *Ss*; over 75% for 26 *Ss*.[9] For 28 of the 32 *Ss*, the accuracy was significant beyond the 5% confidence level; for each of 23, beyond the 1% level.

The accuracy of predictions for the group as a whole was evaluated by two statistical tests: (*a*) Of the total predictions made, 84.6% were correct. (The unweighted mean is 84.7%.) With an N of 32, this deviation from 50% is significant beyond the 1% level. (*b*) Two-thirds or more of the predictions were correct for each of 30 *Ss*. If no systematic factor were operating, this would happen in our group of 32 *Ss* far less than 1 time in 1000.

Accuracy of predictions in various portions of trials.—In a perfectly stable

[9] For one *S*, only 50% of the predictions were correct. This *S* shifted his position frequently during the trials and twice moved so greatly he loosened the lever attached to his eyelid. After the experiment, he reported that he had had an intense recurrent cramp in his leg. Probably it would have been legitimate to drop this *S* who so clearly did not meet the requirements for either theory.

TABLE 6

ACCURACY OF PREDICTIONS FOR EACH QUARTER OF TRIALS AFTER THE FIRST R_+

Quarter of Trials	Responses Predicted Correctly*
First	83.33%
Second	84.33
Third	91.79
Fourth	91.79

* The percentages weight each individual equally. None of the differences between these percentages is statistically significant, each t test giving a $p > .15$, by one-tailed tests of significance.

series of situations, the accuracy of predictions such as those made above should remain constant from one portion of trials to another.

In an unstable series, predictions of this sort should become increasingly accurate as the series progressed. With an unstable series of situations and an increased number of occasions some specified response has been made, more stimuli will have accompanied that response. It thus becomes increasingly likely that a high proportion of any situation will be comprised of stimuli previously present when the response was made. Therefore, when predictions are made by the particular method used in this experiment, their accuracy should increase as the series progresses. (If predictions were made by the more refined use of postremity sketched in the last section, the accuracy should remain constant from one portion of trials to another, even in a very unstable series of situations.)

Since the situations were nearly stable in this experiment, accuracy of our predictions should increase only slightly from the beginning to the end of the series. Table 6 presents the pertinent data.

Further Data

In Fig. 1 and 2 the data for response amplitude and frequency are plotted in the usual way. The shape of these curves strikingly resembles many reported by other investigators. A wealth of experimental data shows that under many sorts of learning conditions, the desired response is made with an increasing frequency and mean amplitude for *groups* of *S*s.

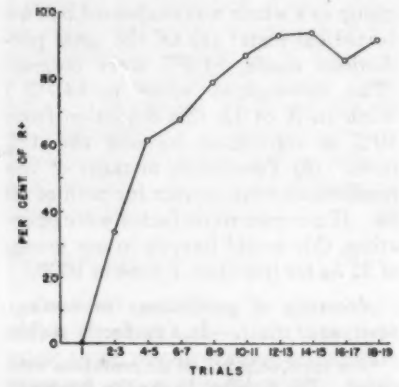

FIG. 1. Probability of R_+ for *S*s grouped. For the first six points, $N = 31$ *S*s; for the last four points, $N = 30.0$, 25.5, 17.5, and 12.5, respectively, which is all *S*s having that many trials or more. The *S* with only nine trials was dropped, so that the first six points would be strictly comparable.

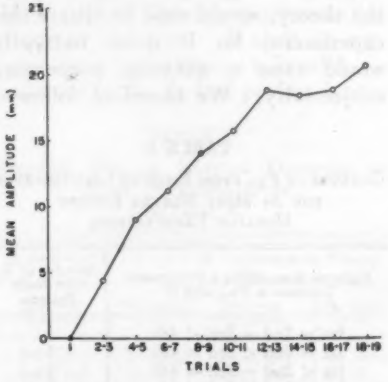

FIG. 2. Mean amplitude of R_+ for *S*s grouped For the first six points, $N = 31$ *S*s; for the last four points, $N = 30.0$, 25.5, 17.5, and 12.5, respectively.

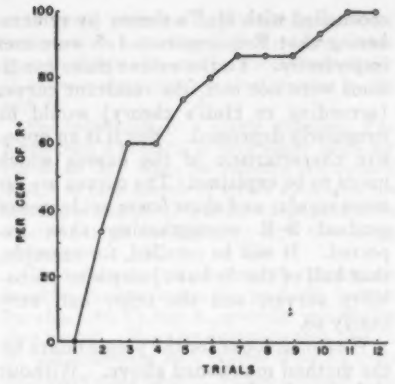

FIG. 3. Probability of R_+ for 15 *S*s each of whom had jumpwise curves

Mainly upon such facts we have based the idea that S–R connections are gradually strengthened with repeated evocations.

But do these facts necessarily mean that an S–R connection is becoming gradually stronger? Although such curves would be expected if there is gradual strengthening of an S–R connection, *they also would be expected if there is not.* This is a matter we sometimes have overlooked.

Suppose different individuals learn at different times. Even if each individual learns suddenly, when the data are grouped the resultant curve will show a progressive increase in P_{R+}. Figure 3 illustrates this. It is a composite for 15 *S*s, each of whom invariably made R_+ after it was once evoked. The group has a gradually increasing P_{R+} strongly resembling that of Fig. 1, but no individual had any gradual increase in P_{R+}. To conclude from such curves that any S–R connection is being gradually strengthened might be a trifle rash.

A similar phenomenon occurs with amplitude. Much of the apparent increase in mean amplitude with N is not due to R_+ becoming larger. Rather, the apparent increase is largely due to different *S*s making their first R_+ on different trials. Figure 4 factors out the effect of this; Fig. 2 does not. Figure 4 shows far less increase in mean amplitude with N.

Under two other conditions such curves as those of Fig. 1 and 2 would result, regardless of whether an S–R connection is gradually strengthened or is established suddenly with full strength. If (*a*) there is much stimulus-variability from one learning occasion to the next or from one test occasion to the next, or if (*b*) there is high similarity between the on-trial stimuli and the between-trial stimuli, the learning curves will increase only gradually.

The first instance confounds the acquisition of several different ${}_sE_R$'s. The greater the stimulus-variability, the greater the variety of stimulus-patterns which must be established as cues for R_+. The smaller the proportion of this total population of stimuli which is present on any specified learning trial, the greater the number of trials needed to establish any specified part of them as cues for R_+. By either Hull's or Guthrie's theory, this will

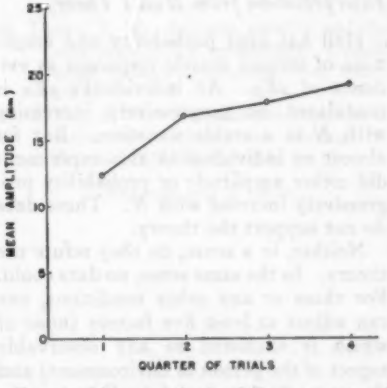

FIG. 4. Mean amplitude of R_+ per quarter of trials following the first R_+

cause probability of R_+ to increase only gradually with N.

We turn now to the second possibility. In terms of Hull's theory, when stimuli which are cues for R_+ subsequently are present when R_+ is made but not reinforced, those S–R connections are weakened. In terms of Guthrie's theory, when stimuli which are cues for R_+ subsequently are present in conjunction with other stimuli which evoke a response incompatible with R_+, those S–R_+ connections are broken. By either theory, the frequency of R_+ would be reduced by the opportunities for unlearning which arise when on-trial stimuli are present between trials; thus P_{R+} would increase more gradually.

The above circumstances would tend to cause also a gradual increase in mean amplitude, since (as generally computed) mean amplitude is a function in part of the number of R_+'s made in each block of trials.

The data of this experiment suggest these three conditions (and not gradual strengthening of S–R connections) may be responsible for the fact that often mean amplitude and frequency gradually increase with N.

Interpretation from Hull's Theory

Hull has used probability and amplitude of striped muscle responses as evidence of ${}_sE_R$. An individual's ${}_sE_R$ is postulated as progressively increasing with N in a stable situation. But for almost no individual in this experiment did either amplitude or probability progressively increase with N. These data do not support the theory.

Neither, in a sense, do they refute the theory. In the same sense, no data could. For these or any other conditions, one can adjust at least five factors (none of which is anchored to any observable aspect of the person or environment) and thus postdict any results whatsoever.

The data of this experiment cannot be reconciled with Hull's theory by remembering that Requirements 1–5 were met imperfectly. To the extent those conditions were not met, the resultant curves (according to Hull's theory) would be irregularly depressed. But it is an opposite characteristic of the curves which needs to be explained: The curves are far more regular and show fewer evidences of gradual S–R strengthening than expected. It will be recalled, for example, that half of the *S*s have jumpwise probability curves, and the other half very nearly so.

We can account for the present data by the method mentioned above. Without violating the theory, we can try different values for ${}_sL_R$ and D, and for the constants in Equations 4 and 29, having various values high for some *S*s and low for others. We also can assume, for some of an *S*'s trials, that ${}_sO_R$ is very high; on others of his trials, very low. By judiciously selecting values for these unanchored symbolic constructs, after careful study of the obtained results, each individual's probability and amplitude data can be postdicted. But by this technique any curve can be postdicted equally well regardless of its shape and *regardless of the conditions under which it was obtained.* (Prediction obviously is not to be expected if we accept Postulate 17 of Hull's most recent theory.)

This experiment deals with one response only. The learning of other S–R connections may be different. Perhaps some S–R connections are established gradually, while S–R's of this type are established suddenly. This remains a logical possibility. However, the S–R connections studied here are of a type previously deemed among the most likely to be gradually acquired.

Summary

An eyelid conditioning procedure was used with 32 human *S*s. Stimulus-variability, similarity of on-trial and off-trial stimuli, and other sources of Hull's various inhibitory potentials were minimized to an unusual extent. The conditions were peculiarly favorable for obtaining smoothly increasing curves.

However, when one largely eliminates those factors which, according to Hull's theory, cause sporadic dips in ${}_{S}E_{R}$, one eliminates also virtually all evidence suggesting a *gradual* increase in ${}_{S}E_{R}$ with N.

For individuals in a stable situation, P_{R_+} jumps from 0 to 100% or very nearly 100% for almost all Ss. When we compare the number of R_+'s in each quarter of an S's trials after his first R_+, in only 16–22% of the comparisons does a later quarter contain more R_+'s than some earlier quarter.

The amplitude data are closer to expectations based on the assumption of a gradually increasing ${}_{S}E_{R}$. An S's first R_+ generally was smaller than his last. However, beginning with his first R_+ and comparing each quarter of an S's trials with every other quarter, we find increases from earlier to later quarters in only 59–66% of the comparisons (not significantly more than 50%).

Predictions of individual responses based largely upon Guthrie's principle of postremity had over 90% accuracy for 15 of the 32 Ss, over 75% accuracy for 26 Ss. Accuracy for most Ss and also for the group as a whole was statistically significant beyond the 1% level.

The data suggest that when P_{R_+} does increase gradually with N, this may not be due to gradual strengthening of any S–R connection. Rather, the gradual increase may occur because (*a*) R_+ must be cued to a variety of situations and not all of the S–R connections are being established on any one trial, (*b*) some unlearning is occurring between test trials, with subsequent relearning being necessitated, and/or (*c*) different Ss learn at different times and their results are being considered together.

It appears quite possible that no S–R is gradually strengthened with repeated reinforcements. Perhaps every S–R is established suddenly, in an all-or-none fashion.

References

1. Beck, L. H. Conditioning and the coordination of movements. *J. gen. Psychol.*, 1939, **20**, 375–397.
2. Bernstein, A. L. Temporal factors in the formation of conditioned eyelid reactions in human subjects. *J. gen. Psychol.*, 1934, **10**, 173–197.
3. Calvin, J. S. Decremental factors in conditioned-response learning. Unpublished doctor's dissertation, Yale Univ., 1939.
4. Cole, L. E. A comparison of the factors of practice and knowledge of experimental procedures in conditioning the eyelid response of human subjects. *J. gen. Psychol.*, 1939, **20**, 349–373.
5. Guthrie, E. R. Conditioning as a principle of learning. *Psychol. Rev.*, 1930, **37**, 412–428.
6. Guthrie, E. R. *The psychology of learning.* New York: Harper, 1935.
7. Hilgard, E. R. Conditioned eyelid reaction to a light stimulus based on the reflex wink to sound. *Psychol. Monogr.*, 1931, **41**, No. 1 (Whole No. 184).
8. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
9. Hull, C. L. *Essentials of behavior.* New Haven: Yale Univ. Press, 1951.
10. Hull, C. L. *A behavior system.* New Haven: Yale Univ. Press, 1952.
11. Marquis, D. G., & Porter, J. M. Differential characteristics of conditioned eyelid responses established by reflex reinforcement. *J. exp. Psychol.*, 1939, **24**, 347–364.
12. Miller, J., & Cole, L. E. The influence of a "voluntary" reaction upon the development and the extinction of the conditioned eyelid reaction. *J. genet. Psychol.*, 1936, **48**, 405–439.
13. Porter, J. P. Temporal factors in experimental extinction. Unpublished doctor's dissertation, Yale Univ., 1936.
14. Switzer, S. A. Backward conditioning of the lid reflex. *J. exp. Psychol.*, 1930, **13**, 76–97.
15. Voeks, V. W. Stimulus stability and the acquisition of S–R connections. Unpublished doctor's dissertation, Yale Univ., 1947.
16. Voeks, V. W. Formalization and clarification of a theory of learning. *J. Psychol.*, 1950, **30**, 341–363.

(Received July 2, 1953)

Journal of Experimental Psychology
Vol. 47, No. 3, 1954

VISUAL DISTANCE DISCRIMINATION IN THE RAT[1]

ANN GREENHUT

Armament Systems Training Research Laboratory, Human Resources Research Center

In order to investigate the visual perception of distance in infrahuman *Ss* it is necessary to choose a criterion of behavior on the basis of which it may be inferred such that *S* is able to make adequate responses only by using visual cues which arise from the relative distance of the stimulus object. One of the common pitfalls of experiments on visual distance discrimination is that *S* has been allowed to use kinesthetic as well as visual cues. Some aspect of the jumping of the animal has been most often selected as a criterion, such as latency (**12, 13**), accuracy (**7**), and force (**1, 8**) of the jump. These studies have been reviewed in an earlier publication (**1**). In essence, there has been considerable difficulty in using this particular response of rats, and inconsistent results have been reported. Hence, it was believed that a criterion which did not require *S* to jump the distances presented would have definite advantages.

In the present study a modified form of the Howard-Dolman test (**4**) has been adopted, thus using a criterion which has never been applied to infrahuman animals. The criterion of the presence or absence of visual distance discrimination was whether or not rats could learn to discriminate between two pegs at different distances from them by going to the nearer peg when all other cues were controlled. The psychophysical method of constant stimulus differences with different distance standards was used to determine the presence and precision of such discrimination in albino and pigmented rats.

[1] This article is based on a dissertation submitted to the Graduate School of the State College of Washington in partial fulfillment of the requirements for the Ph.D. degree. The author wishes to thank Dr. F. A. Young, who directed this research.

METHOD

Apparatus.—The floor plan of the apparatus is shown in Fig. 1. The pegs (shown at a typical setting) could be moved varying distances from the choice point (at the entrance of the alleys) by adjusting the middle, movable portions of the floor to which they were attached. The entire maze was painted a flat white except for the black pegs used as stimulus objects, the unpainted hardboard floor, and the black goal box. The alleys and the choice box were covered with glass; the remainder of the maze, with wire netting. The experimental room had burlap-covered walls to eliminate glare and to provide a uniform background. Lighting was both uniform and indirect.

Subjects.—Twenty-four female rats (12 albinos of the Wistar strain and 12 black hooded) were used. Three of the latter had to be discarded at various stages of the experiment because of refusal to run. All the rats were approximately 90 days old at the beginning of the experiment. The only motivation specifically employed was hunger.

Procedure.—The procedure consisted of three parts: preliminary training (initial and transitional phases), test, and transposition series. In the initial phase of training, *Ss* were rewarded for going to the nearer of two pegs when one peg was placed directly by the choice point within reach of *S* while the other was placed at the far end of the second alley. The goal box containing a pellet of food was placed inside the end box of the alley with the nearer peg. Each *S* received ten trials a day, the five trials with each alley positive being presented in a random order. The criterion of learning was 90% correct in 20 trials. In the transitional phase of training, the nearer peg was moved gradually away from the choice point and the far peg toward the choice point. The following pairs of distances from the choice point (in inches) were successively presented: 3–27, 4–24, 6–22, 7–20, 8–22, 10–24. Each setting was presented until 85% of the trials were run correctly on two successive days, except in the final setting in which the 90% criterion was used.

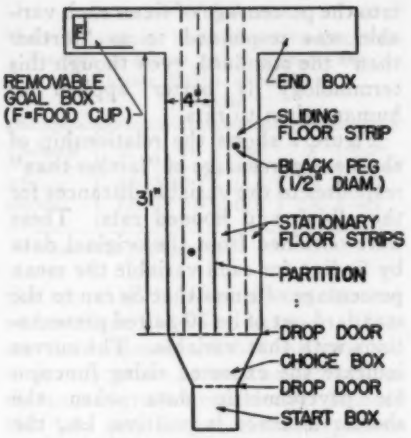

FIG. 1. Floor plan of apparatus

It may seem, prima facie, that visual distance perception has been demonstrated in the transitional phase since all of the rats were able to learn to choose the 10-in. distance over the 24-in. distance 90% of the time on two successive days. However, before drawing this conclusion there are at least two considerations which must be taken into account. First, other cues must have been fully controlled. If there were any chance, for example, for *S* to use kinesthetic cues before making its choice, as was the case in Takatsuki's study (9), then the inference would be unjustified. In the present investigation the rat had to make its choice *before* traversing the alley. Moreover, the side on which the nearer peg appeared was randomly determined and, consequently, it was not possible for *S* to make direct use of its previous runs in making its choice.

Brightness was controlled by an evenly lighted maze and by using small, flat black pegs which would not reflect appreciably different amounts of light from the distances used. It is highly unlikely that there were any distinctive food odor cues. The *E* always stood directly behind *S* in order to avoid giving any cues inadvertently. With these precautions the discrimination of the 10-in. peg from the 24-in. peg must have been made on a visual basis.

Secondly, the conclusion presupposes that *S* was discriminating on the basis of relative distance alone. To determine whether the maximum distance used was in the range of vision, four albino *S*s were trained to choose the alley in which there was a peg as against an alley in which there was none. All were able to make successful responses when the peg was at least 26 in. away. During the preliminary phase of the experiment, however, while both pegs were within the range of vision of *S*s, their responses could have been made on the basis of the absolute distance to the nearer peg alone, and hence would give no definite ground for inferring that visual *distance* discrimination was involved.

In order to necessitate that *S* respond to the relative position of the pegs, a series of separations was presented in which a peg 10 in. from the choice point was paired with pegs which were both closer to and farther from it, thus making the 10-in. distance positive only half the time. The distances first paired with 10 in. were always at one extreme of the alley (2 in. or 25 in. from the choice point).

The test series was then run to determine whether *S* could make visual distance discriminations and to obtain an estimate of the difference threshold. The nearer peg was made positive in the same manner as in the preliminary training. The method of constant stimulus differences was used with 10 in. as the standard against which ten variable distances were paired. The variables were presented in a random order. In the present case *E* used a lottery technique without replacement. The order thus obtained was used twice, one time so that the given variable would appear on the left and the next time so that it would appear on the right. This, of course, was a precaution against the space error.

On the basis of an earlier, exploratory study, a 2-in. increment between the variables was selected for the albinos. The variables ranged from 4 in. to 24 in. In order to avoid the possibility that the discrimination might break down, the 10-in. distance was not used as a variable. It was assumed that 50% of such trials would have been run to the standard. It was necessary to use a smaller increment for the pigmented rats than for the albinos. A 1-in. interval between the variables was selected, with the distances ranging from 6 in. to 16 in. Each variable was presented 20 times with the standard, making a total of 200 trials for each *S*.

In the final or transposition phase, 15-in. and 20-in. standards were used successively with the albino group, except for one *S* which was discarded for refusal to run. Five variables which were paired with the 15-in. standard ranged from 3 in. to 23 in., separated by increments of 4 in., while five variables paired with the 20-in. standard ranged from 5 in. to 30 in. with increments of 5 in. Ten trials were run for each variable. In no case was the standard paired with itself.

RESULTS AND DISCUSSION

The Howard-Dolman test to be used with infrahuman *S*s requires first, of course, that the animal be trained to respond to a certain type of situation,

TABLE 1

MEANS AND *SD*'S ON NUMBER OF DAYS REQUIRED IN PRELIMINARY TRAINING*

Group	*N*	Initial Phase		Transitional Phase**	
		M	*SD*	M	*SD*
Albinos	12	18.6	3.1	26.9	5.7
Hooded (retained)	9	16.0	2.4	25.1	3.8
Hooded (discarded)	3	21.7	3.7	27.0	3.6

* Ten trials were run each day.
** See text for separations used.

viz., the nearness and farness of the rods. The nature of the test was slightly modified in so far as the animal was able to see and to traverse the floor of the maze to which the rods were attached, whereas in the Howard-Dolman apparatus only the middle portions of the pegs are shown. The mean number of days required for the learning of this initial phase is shown in Table 1 for both albino and hooded *Ss*. The data on the discarded hooded *Ss* are given separately. In comparison with the "retained" hooded *Ss*, they took significantly more days to learn initially ($p < .05$) but did not differ significantly in mean number of correct trials per day during initial training nor in any respect on the transitional phase.

Presence of visual distance discrimination.—Data as those obtained in the test series by the method of constant stimulus differences are commonly treated in terms of the percentage of times each variable was responded to as heavier than, brighter than, or larger than the standard, depending upon the kind of stimulus studied. The advantages of using this treatment are that the judgments may be evaluated in terms of the expected ogival function (the phi-gamma hypothesis) and it is possible to estimate the difference threshold. Accordingly, the present data have been converted into the percentage of times each variable was responded to as "farther than" the standard, even though this terminology is better applied to humans than to rats.

Figure 2 shows the relationship of the mean percentage of "farther than" responses to the variable distances for the albino and hooded rats. These were obtained from the original data by finding for each variable the mean percentage of times that *Ss* ran to the standard out of its 20 paired presentations with that variable. The curves indicate the expected rising function for psychometric data when the shorter distance is positive, i.e., the larger the variable the more frequently the standard was chosen. If the rats had not been discriminating, one would expect a roughly horizontal line through the 50% point of the graph, i.e., each response would have been merely a chance selection. Since the curves do show the expected rising function, it is inferred that the rats were making visual distance discriminations.

It should be emphasized that in the test phase the variables were randomized in order of presentation. By such an arrangement it was necessary for

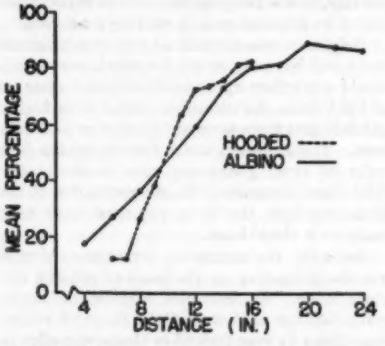

FIG. 2. Group curves showing the percentage of "farther than" responses for each variable with 10 in. as standard. The 10-in. value has not been plotted since it was not obtained empirically but is assumed to be 50%.

TABLE 2

MEAN QUARTILES AND SEMI-INTERQUARTILE RANGES IN INCHES FOR 9 HOODED AND 12 ALBINO RATES WITH 10-IN. STANDARD

Subjects	Q_1	Q_3	Q_3	Q	SD*
Albinos	6.0	10.3	15.6	4.8	7.3
Hooded	7.9	9.9	13.4	2.8	4.1

* Obtained by converting Q.

the animals to respond to relative rather than absolute distance. The standard represented the positive stimulus only when it was paired with a distance of greater magnitude. Hence, the only way for successful responses to be made consistently (better than chance) was for the animal to compare the two stimuli.

Discriminal sensitivity of hooded and albino rats.—Accepting, then, the above interpretation, the precision of this type of discrimination remained to be determined. Several statistical methods are available for handling the data in this form. By the graphic method, which was found most satisfactory, the various quartiles may be obtained. In the present case, these indicate the variable stimuli which were responded to as "farther than" the standard for the given percentage of times. All the individual curves suggested an ogival function although there were some irregularities. The semi-interquartile range, Q, can be used as an indication of the sensitivity of S. The smaller Q is, the greater the sensitivity. Referring to Table 2, the mean Q is 4.8 in. for the albinos and 2.8 in. for the hooded rats. The semi-interquartile range covers 25% of the cases on either side of the median if the distribution is symmetrical. Psychometric data are not strictly symmetrical, as Thurstone (10) has shown, but in so far as symmetry is approached it may be said that the middle 50% of the "farther than" responses should generally fall between 5.5 in. and 15.1 in. for the albinos, and between 7.1 in. and 12.7 in. for the hooded rats. The Q values may be converted to standard deviations, which are also presented in Table 2. The difference between mean standard deviations for the hooded and albino groups is significant ($t = 7.2$) beyond the .01 level of confidence, indicating that the hooded animals evidence a greater sensitivity than the albinos. This is in agreement with Russell (7) and Greenhut (1).

It is true that hooded animals were presented with smaller intervals of separation between variables than were the albinos. However, this difference in itself indicates greater sensitivity of the former group. Had smaller intervals been unjustified, this would have been indicated by a greater dispersion of errors or larger semi-interquartile ranges, but on the average the hooded group made far fewer errors than the albinos, significant beyond the .01 level of confidence. This is in accord with results from other studies, e.g., Hirsch and Weymouth's (3) finding that a positive relationship exists between visual acuity and depth perception, and Lashley's (5) finding that

TABLE 3

MEAN PERCENTAGES OF "FARTHER THAN" RESPONSES FOR EACH VARIABLE STIMULUS WITH 15- AND 20-IN. STANDARD STIMULI

(N = 11 albino rats)

15-in. Standard		20-in. Standard	
Variable Stimulus	Mean %	Variable Stimulus	Mean %
3	8.2	5	6.4
7	11.8	10	11.8
11	25.5	15	20.0
15	50.0*	20	50.0*
19	62.7	25	61.8
23	86.4	30	76.4

* The standard was not paired with itself, the 50% presented being assumed.

TABLE 4

MEAN QUARTILES AND SEMI-INTERQUARTILE RANGES IN INCHES FOR 15- AND 20-IN. STANDARD STIMULI
(N = 11 albino rats)

Standard Stimulus	Q_1	Q_3	Q_2	Q	SD*
15-in.	9.9	15.2	21.5	5.8	8.6
20-in.	14.6	21.7	29.2	7.3	10.8

* Obtained by converting Q.

pigmented rats possess greater visual acuity than albino rats.

Tests with other standards.—Only the albino group was used for succeeding tests, which may be called generalization tests. In Table 3 are presented the mean percentages of "farther than" responses for each variable against 15-in. and 20-in. standards. As in the case with the 10-in. standard, the percentages show the consistent increase with variable distance which is expected for psychometric data. The curves (not presented) suggest an ogival function in which the final negatively accelerated part must be extrapolated, i.e., larger variable distances should have been included to obtain the complete curve.

By the graphic method the mean quartiles and semi-interquartile ranges were determined for each standard. These are shown in Table 4. The Q_3 value shows that on the average 21.5 in. was just noticeably differentiated from 15 in., and 29.2 in. from 20 in. These data obtained with the 15- and 20-in. standards are of value in indicating the occurrence of stimulus generalization.

SUMMARY

A modification of the Howard-Dolman or rod test of visual distance discrimination was employed with 12 hooded and 12 albino rats. The *S*s were initially trained to choose the alley which contained the nearer of two pegs, food being the reward. The test situation used the method of constant stimulus differences with a standard of 10 in. and ten variable distances. The results show a positive relation between the magnitude of the variables and percentages of "farther than" responses. It was inferred that the rats were making visual distance discriminations, since the only way successful responses could consistently be made was through the cue given by the relative distance of the pegs. Kinesthetic, tactual, and brightness cues were controlled. The hooded rats showed greater sensitivity to relative distance than the albino rats. Further tests with the albino rats indicated that in this mode of discrimination stimulus generalization occurs.

REFERENCES

1. GREENHUT, A., & YOUNG, F. A. Visual depth perception in the rat. *J. genet. Psychol.*, 1953, **82**, 155–182.
2. GUILFORD, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
3. HIRSCH, M. J., & WEYMOUTH, P. Distance discrimination: VI. Relationship of visual acuity to distance discrimination. *J. Aviat. Med.*, 1948, **19**, 56–58.
4. HOWARD, H. J. A test for the judgment of distance. *Amer. J. Ophthal.*, 1919, **2**, 656–675.
5. LASHLEY, K. S. The mechanism of vision: III. The comparative visual acuity of pigmented and albino rats. *J. genet. Psychol.*, 1930, **37**, 481–484.
6. MUNN, N. L. *Handbook of psychological research on the rat.* New York: Houghton Mifflin, 1950.
7. RICHARDSON, F. A study of sensory control in the rat. *Psychol. Monogr.*, 1909, **12**, No. 1 (Whole No. 48).
8. RUSSELL, J. T. Depth discrimination in the rat. *J. genet. Psychol.*, 1932, **40**, 136–159.
9. TAKATSUKI, R. Vision as a factor of distance discrimination in the albino rat. *Acta Psychol. Keijo*, 1937, **3**, 1–32.
10. THURSTONE, L. L. The phi-gamma hypothesis. *J. exp. Psychol.*, 1928, **11**, 293–305.
11. WASHBURN, M. F. *The animal mind.* New York: Macmillan, 1926.
12. WAUGH, K. T. The role of vision in the mental life of the mouse. *J. comp. Neurol. Psychol.*, 1910, **20**, 549–599.
13. YERKES, R. M. Space perception of tortoises. *J. comp. Neurol. Psychol.*, 1904, **14**, 17–30.

(Received June 6, 1953)

Journal of Experimental Psychology
Vol. 47, No. 3, 1954

TRANSFER IN VERBAL MATERIALS WITH DISSIMILAR STIMULI AND RESPONSE SIMILARITY VARIED

ROBERT K. YOUNG AND BENTON J. UNDERWOOD

Northwestern University

It is a common generalization (e.g., 4, 6) that positive transfer occurs when stimuli are different and responses are similar or identical. This paradigm (A–B, C–B) is said to produce positive transfer, while A–B, A–C is said to produce negative transfer. That positive transfer occurs between two verbal tasks having the A–B, C–B relationship is supported by the work of Bruce (1). Furthermore, since Bruce found little or no positive transfer when both stimuli and responses of the two lists were dissimilar, it is inferred that when stimuli are different, positive transfer increases as response similarity increases. However, a theoretical basis for such a principle has never been adequately explored. The purpose of the present paper is to advance and test a hypothesis concerning the processes involved in transfer when stimuli are different and response similarity varied.

On the basis of a theory of stimulus and response generalization (7, 10) it has been shown that if stimuli in two paired-associate lists are dissimilar, variation in response similarity should produce no changes in amount of positive transfer. While response similarity has been shown to be a major factor influencing positive transfer (11), such similarity cannot influence transfer (theoretically) unless a certain level of stimulus similarity is present. If, then, positive transfer does increase as interlist response similarity increases when stimuli are dissimilar (as suggested by Bruce's work), transfer theories based on direct stimulus and response generalization are incomplete. We propose that the missing portion of the theory required to account for Bruce's findings is *transfer of response differentiation.*

Gibson (2), in her theory of transfer, proposes that in the A–B, A–C paradigm, one source of positive transfer lies in the transfer of stimulus differentiation. By this she means that, in so far as the stimuli *within* a list are similar, a part of the learning process involves the establishment of differentiation among these stimuli. Presumably, such differentiation would be established in learning the first list and would transfer to the learning of the second (where the stimuli are identical to those in the first), thus reducing somewhat the time to learn the second list. The present proposal assumes that exactly the same process may take place for responses in the A–B, C–B paradigm. Transfer of response differentiation is assumed to account for some of the positive transfer which occurs. Furthermore, it can be seen that as the responses between lists are made less and less similar, transfer of response differentiation would become less and less. Thus, we would predict that as responses between the lists become less and less similar, positive transfer would become less and less until no transfer should occur (owing to transfer of response differentiation) when the responses between the two lists are quite dissimilar.

Transfer of response differentiation from one list to another should be minimal when *intralist* response sim-

ilarity is low. That is, with low intralist response similarity, the establishment of differentiation among responses within a list will be a very minor component of the learning process; hence, there will be little differentiation to transfer. Having low intralist response similarity, therefore, would be one way in which transfer as a function of interlist response similarity between lists (with stimuli dissimilar) could be minimized. Another technique by which transfer of differentiation could be reduced would be to predifferentiate the responses before the two lists used to study transfer are learned. Such predifferentiation would eliminate the necessity of intralist response differentiation in learning the first list and reduce the apparent positive transfer in learning the second list.

In the present study we have used both techniques mentioned above. In the first place we have made intralist response similarity as low as careful inspection procedures allow. Since we feel that it is virtually impossible to remove all similarity from among a series of ten adjectives, we might still expect differential positive transfer as a function of response similarity between lists. Therefore, we have added predifferentiation of responses as a technique to insure that there will be no transfer of differentiation from the first to second list in our transfer tests. It is clear that under these conditions our theory must predict that there will be no differential transfer as a function of interlist response similarity when the stimuli in the two lists have both low interlist and intralist similarity.

Obviously the above proposal does not predict that positive transfer will be eliminated. The theory predicts only that there will be no relationship between transfer and interlist response similarity when predifferentiation among responses has been carried out. Sources of positive transfer which are presumably not related to specific interlist response similarity (such as learning-how-to-learn) will still be operative.

Procedure

General procedure.—Two parallel experiments were carried out. These two experiments differ in only one way: in one experiment the responses used in the transfer lists were predifferentiated, while in the other they were not. The group of *Ss* serving in the first experiment will be referred to as the PD Group (predifferentiation group); *Ss* in the second group will be called the IR Group (irrelevant predifferentiation group). The *Ss* in this second group were given the same type of pretask to practice as *Ss* in the PD Group, but items in this pretask were irrelevant to the transfer tasks. Such a control is necessary to keep both groups at the same practice level. According to the theory, the PD Group should show no relationship between interlist response similarity and transfer. The IR Group may show such a relationship and if so it is to be attributed to intralist response similarity which, as mentioned before, was low, but which could still be sufficient to produce some transfer of response differentiation.

Within each experiment there were three experimental conditions, each condition representing a different level of interlist response similarity. Each condition in turn consisted of the learning of two lists of paired adjectives after the predifferentiation training. It was from the paired-adjective lists, of course, that transfer as a function of interlist response similarity was measured.

Paired-adjective lists.—The lists used to study transfer consisted of ten pairs of two-syllable adjectives. One set of two lists had low interlist response similarity; a second set had medium interlist response similarity (as determined by Haagen, 3); and a third set had identical responses (high similarity). In all three sets the interlist- and intralist-stimulus similarity was as low as could be achieved by careful inspection of the adjectives. Thus, we have three sets of two lists each, which are presumed to vary only in interlist response similarity. Lists were learned by the anticipation method from a Hull-type drum with a 2-sec. exposure of the stimulus and 2 sec. for the stimulus and response together. Time between successive trials was 4 sec. Three different orders of presentation were used to mini-

mize serial associations. Each list was learned to a criterion of one perfect trial.

Predifferentiation lists.—On each experimental day *S*s had predifferentiation training before learning the two adjective lists. Such training was carried out by the verbal-discrimination technique. By this method *S* is presented a series of pairs of adjectives with instructions that one of each pair has been arbitrarily selected as the correct response. Each pair was shown for 2 sec. (one above the other), and then the shutter on the drum was lifted to expose the correct member of the pair for 2 sec. On successive trials the spatial positions of the two words of a pair were randomized to make position an inconsistent cue. Learning of the discrimination lists was always carried to one perfect trial.

The nature of the words used in the predifferentiation lists is critical, of course, if differentiation among responses used in the adjective lists is to be established. In the two paired-associate lists for the low-similarity condition there is a total of 20 responses. Preceding the learning of these lists, therefore, *S*s in the PD Group learned a verbal-discrimination list made up of the 20 responses in the paired-associate lists, these 20 responses being presented as 10 pairs in the verbal-discrimination task. A response item from one of the paired-associate lists was paired at random with a response item from the other. Half of the words from one list were correct in verbal-discrimination learning and half were incorrect. Thus, no bias would be attached to learning either the first or second paired-associate lists as a consequence of differential frequency of approach and avoidance of responses in those lists during verbal-discrimination learning.

Preceding the learning of the two adjective lists with medium interlist response similarity, *S*s in the PD Group likewise learned a verbal-discrimination list made up of the 20 responses in the two paired-associate lists. In constructing this list the similar items were paired. All other details were the same as for the low-similarity condition as described above.

It will be remembered that high interlist response similarity between the paired-associate lists consisted of identical responses. Therefore, in order to construct verbal-discrimination lists of ten pairs it was necessary to add ten additional items. These ten words were chosen from Haagen's list (3) so that each was highly similar to a response in the paired-associate lists. Again, these similar items were paired for the verbal-discrimination learning.

The verbal-discrimination learning described above was used for the PD Group. For the IR Group it was necessary only to give practice in verbal-discrimination learning so that practice level and any warm-up effects would be equal for the two groups. For this purpose, additional verbal-discrimination lists were constructed using words from Melton's (5) materials. The words were all two-syllable adjectives, but in no case was there any apparent similarity between these words and those which *S*s subsequently were presented with in the paired-adjective lists. The IR Group, therefore, should be equal in practice and warm-up with the PD Group but should have no transfer of differentiation from the verbal-discrimination lists to the paired-adjective lists as was presumed to be the case in the PD Group.

Practice day.—All *S*s were given one practice day in which one verbal-discrimination list and two paired-adjective lists were learned. The relevance of the verbal-discrimination list to the paired-associate lists which followed and the degree of response similarity between adjective lists were a composite of all conditions in both the PD and IR experiments. That is, some interlist responses were similar, others were dissimilar; some verbal-discrimination items were relevant to the learning to follow while others were irrelevant. The two groups of *S*s were matched on the basis of their performance on the first paired-associate list learned on the practice day.

Design details.—Each group of *S*s consisted of 18 undergraduate students. Each *S* served in all three experimental conditions of his experiment, with the order of conditions counterbalanced. The conditions, of course, are the three sets of lists which differed in interlist response similarity. Furthermore, the two lists within a set were counterbalanced so that half of the time one served as the first list and half of the time the other served as the first list. List differences within a set can, thereby, be balanced out and transfer effects can be determined by comparing performance on the first-learned list with that of the second without regard to the specific list involved.

On each experimental day *S* first learned a verbal-discrimination list followed by two paired-adjective lists. The time between the verbal-discrimination list and the first paired-adjective list was approximately 60 sec., and between the two paired-adjective lists the interval was approximately 30 sec. In most cases the four sessions (one practice and three experimental) were run on four consecutive days. Most *S*s had previously served in experiments in which nonsense syllables had been used.

Results

Equality of groups.—The two groups were matched in learning the first

TABLE 1

ANALYSIS OF VARIANCE OF DIFFERENCE SCORES BASED ON NUMBER OF TRIALS TO ATTAIN ONE PERFECT RECITATION

Source	*df*	Mean Square	*F**
Groups	1	8.882	
Subjects/Groups	34	33.663	
Similarity	2	57.149	1.93
Similarity × Groups	2	36.045	1.21
Days	2	53.482	1.80
Days × Groups	2	25.489	
Residual/Groups	64	29.683	
Total	107		

* With 2 and 64 *df*, *F* of 3.14 is needed for the .05 level of confidence.

paired-associate list on the practice day. The mean number of trials to learn this list by the PD Group was 14.22 ± 1.66, and for the IR Group, 14.50 ± 1.58. It is apparent that these two means are quite comparable. The product-moment correlation between number of trials to learn this practice list and the number of trials to learn the first paired-associate list for all experimental conditions combined was .70 ± .09.

Transfer as measured by trials to learn.—Transfer effects may be determined by examining differences in number of trials to learn to one errorless repetition under each condition. These differences will be called difference scores. Such scores must be used if an adequate test of the theory is obtained, for the effect of predifferentiation, according to the theory, is to reduce the number of trials to learn the first list. The effect of this is to reduce the apparent transfer from the first to the second list. Both groups may learn the second paired-associate list for a given condition equally rapidly but the amount of transfer may differ when learning of the first list is used as a base.

Analysis has shown that the three sets of paired-associate lists were not equal in difficulty. This fact does not prohibit comparisons between groups and only means that for comparisons within groups the difference scores must be used.

The analysis of variance for the six distributions of difference scores (three levels of response similarity for each group) is shown in Table 1. Variance has been shown not to be heterogeneous. Only *F*'s greater than unity have been entered in the table. For testing the difference between groups the Subjects/Groups term is appropriate. All other sources are evaluated by Residual/Groups.

It can be seen that none of the effects of the variables approaches statistical significance. Thus, two conclusions can be reached: (*a*) the two groups did not differ in the amount of transfer, and (*b*) transfer did not vary as a function of interlist response similarity.

In Fig. 1 we have plotted the percentage transfer for each group for each level of response similarity. Percentage transfer was calculated by dividing the difference in number of trials to learn the first and second lists by the number of trials to learn the

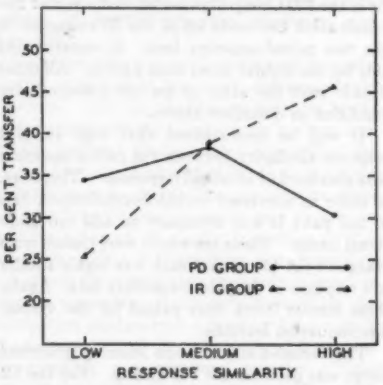

FIG. 1. Percentage positive transfer for both groups for three levels of response similarity

first. As indicated above, neither curve can be considered statistically different from a horizontal line. This means that even without predifferentiation of responses, transfer does not vary as a function of interlist response similarity when stimuli are dissimilar. According to our theory, this should be the case if intralist response similarity is very low. Nevertheless, Fig. 1 shows trends which indicate that predifferentiation had a small but expected effect. As response similarity increases, positive transfer increases for the IR Group. This is due, according to the theory, to the transfer of response differentiation from the first to the second list, this transfer being greatest when the responses in the two lists are identical. With predifferentiation of responses (PD Group) no such trend is noticeable and transfer is clearly not related to interlist response similarity.

One final fact should be noted in Fig. 1. While transfer is not significantly related to interlist response similarity, it is clear that large amounts of positive transfer were present in all conditions. All points differ from zero at better than the 5% level of confidence.

Other analyses.—Other analyses of the data which have been made will be mentioned briefly, since none adds any new information. We have compared the differences in number of correct responses given on the first two trials of the first list and the first two trials of the second list. These difference scores showed more transfer for the IR Group than for the PD Group for conditions of medium and high interlist response similarity, but these differences were not significant statistically. Thus, the same trends are shown in these scores as in those shown in Fig. 1.

If predifferentiation is effective we would expect that the number of trials to learn the first paired-adjective list for a given condition would be less for the PD Group than for the IR Group. In only two of the three conditions was this the case, and again these differences were not significant. Likewise, if predifferentiation is effective we would expect fewer errors in learning the first paired-adjective list for the PD Group than for the IR Group. In two of the three conditions the differences were in the expected direction but again no statistical significance can be attached to them.

In summary it can be seen that if the predifferentiation training had any effect, it cannot be shown statistically.

Discussion

The results of the present study show that positive transfer in verbal materials does not vary as a function of interlist response similarity when interlist stimulus similarity is low. The implication is that A–B, C–B cannot be thought of as a basic paradigm of positive transfer; any other relationship between responses in the two lists, including complete dissimilarity, is just as much a basic paradigm.

Trends in our data have shown that positive transfer will vary as a function of interlist response similarity when interlist stimulus similarity is low. According to the present formulation this may be attributed to the transfer of response differentiation from the first to the second list. If lists were constructed with considerable intralist response similarity we would expect these trends to become statistically significant. It will be remembered that our lists were intentionally constructed to have low intralist response similarity in an effort to show that even without predifferentiation no differential transfer would result as a function of interlist response similarity. Statistically speaking, this attempt was successful. We must conclude, therefore, that when intralist response similarity is low, interlist response similarity is not related to

amount of transfer when interlist stimulus similarity is low.

Previous results (1) showing that interlist response similarity was an effective variable in this situation may be attributed to the transfer of response differentiation from the first to the second list, this transfer being greater the greater the interlist response similarity. We suspect that Bruce's lists had enough intralist response similarity to make transfer of differentiation of these responses appreciable. According to our conception, the maximum amount of transfer of such differentiation will occur when both intralist and interlist response similarity are high.

While our data do not show that interlist response similarity is an effective variable, they do show that all conditions produced significant amounts of positive transfer. Certain processes which may contribute to this transfer will be suggested.

In our experiment learning-how-to-learn was not, statistically, a significant variable. This probably resulted from the fact that most *S*s had served in previous experiments. Nevertheless, there was approximately a 15% increase in rate of learning the adjective lists on the last experimental day as compared with the first. Therefore, a small amount of the total transfer between two lists within a set may be attributed to learning-how-to-learn.

Other experiments (e.g., 11) have shown that the higher the degree of first-list learning the greater the positive transfer. Even in an experiment which produces negative transfer on the initial trials of learning paired-adjective lists, the transfer effect may be positive when transfer measurements are made in terms of a criterion of one perfect trial (8). Therefore, the learning of the first list to one perfect trial may account for some of the positive transfer obtained, although we do not know that this variable (degree of learning) operates independently of certain similarity relationships.

Lack of warm-up before learning a first list in a transfer situation may result in apparent positive transfer. This comes about because the first list serves as a warm-up task for the second; consequently, the second list is learned more rapidly than the first because of this warm-up. Other research recently completed (unpublished) indicates that verbal discrimination does serve as a warm-up task for paired-associate learning, but we do not know that this warm-up is as complete or fully effective as would be the learning of a paired-associate list. Incomplete warm-up would tend, therefore, to increase the measured positive transfer.

The above factors may be invoked to account for the positive transfer observed in all conditions of the present experiments. It is clear, however, that we have no way of telling from our present results whether they will account for the full amount of measured transfer. Other processes, as yet unidentified, may be operative.

Summary

The A–B, C–B paradigm in verbal learning is commonly said to be the basic paradigm for positive transfer. Furthermore, it is commonly believed that as the responses in the paradigm become less and less similar, positive transfer decreases. Such a relationship cannot be predicted strictly on the basis of interlist stimulus or response generalization. However, such a relationship would be expected if there is a transfer of response differentiation from the first to the second task.

The present experiments studied positive transfer in paired-associate lists when interlist response similarity was varied from identity to low similarity, with the interlist stimulus similarity always low. Lists were constructed so that intralist response similarity was low. Under such conditions generalization theory would predict no relationship between response similarity and transfer. However, since it is difficult to remove all similarities among responses, two parallel experiments were performed in which one provided for predifferentiation of responses before learning the first paired-associate list and the other did not. If transfer of response differentiation is effective even when responses have low intralist similarity, the predifferentiation procedure would eliminate it from transfer measurements.

Predifferentiation of responses was carried out by verbal-discrimination learning prior to the learning of the two paired-associate lists for a

given condition of response similarity. Learning of all lists was carried to one perfect trial.

The results show: (*a*) Interlist response similarity was not related significantly to amount of transfer in either experiment. (*b*) The predifferentiation procedure had a small but statistically insignificant effect on transfer. (*c*) Significant positive transfer occurred in all conditions.

These results indicate that A–B, C–B cannot be considered a basic paradigm of positive transfer any more than can a paradigm in which responses are quite dissimilar. Whether or not interlist response similarity is related to amount of positive transfer (when stimuli are dissimilar) probably depends upon the level of intratask response similarity.

References

1. Bruce, R. W. Conditions of transfer of training. *J. exp. Psychol.*, 1933, **16**, 343–361.
2. Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.
3. Haagen, C. H. Synonymity, vividness, familiarity, and association-value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, **30**, 185–200.
4. McGeoch, J. A., & Irion, A. L. *The psychology of human learning*. New York: Longmans, Green, 1952.
5. Melton, A. W. Materials for use in experimental studies of learning and retention of verbal habits. Mimeo. manuscript, Univ. of Missouri, 1940.
6. Munn, N. L. *Psychology*. (2nd Ed.) Boston: Houghton Mifflin, 1951.
7. Osgood, C. E. The similarity paradox in human learning. *Psychol. Rev.*, 1949, **56**, 132–143.
8. Underwood, B. J. Retroactive and proactive inhibition after five and forty-eight hours. *J. exp. Psychol.*, 1948, **38**, 29–38.
9. Underwood, B. J. Proactive inhibition as a function of time and degree of prior learning. *J. exp. Psychol.*, 1949, **39**, 24–34.
10. Underwood, B. J. *Experimental psychology*. New York: Appleton-Century-Crofts, 1949.
11. Underwood, B. J. Associative transfer in verbal learning as a function of response similarity and degree of first-list learning. *J. exp. Psychol.*, 1951, **42**, 44–53.

(Received June 18, 1953)

Journal of Experimental Psychology
Vol. 47, No. 3, 1954

THE ROLE OF OVERT ERRORS IN SERIAL ROTE LEARNING

HELEN SCHEIBLE AND BENTON J. UNDERWOOD

Northwestern University

In a previous paper (4) it was shown that differences in frequency of overt errors during serial learning of nonsense syllables were not associated with differences in rate of learning. The differences in error frequency resulted from different rest-interval activities introduced during distributed learning trials. In that same paper it was also shown that for either of the two rest-interval activities used, the correlations between mean errors per trial and trials to learn were not significantly different from zero. Thus, the evidence suggests that frequency of overt errors in serial learning neither influences nor is associated with rate of learning. We feel that establishing the generality of such a conclusion is important for two reasons: First, we suspect that such a conclusion runs counter to a commonsense conception of learning. Second, such a conclusion would indicate that differential reinforcement takes place as rapidly without overt errors as with overt errors.

The present study represents an attempt to influence greatly the frequency of overt errors in serial learning by varying instructions to the *S*. If such differences in instructions result in widely different error frequencies, and if rate of learning remains unrelated to error frequency, the generality of the previous evidence is broadened.

Procedure

General design.—Three groups, each composed of 24 college students, served as *S*s. Each group served in three experimental sessions, learning a different list of nonsense syllables at each session. The conditions for the three groups differed only in the instructions concerning the making of overt errors. Some previous evidence (4) had suggested that if there is a relationship between overt error frequency and rate of learning, it is dependent upon the stage of practice. For that reason, the three sessions were used so that the relationship between errors and rate of learning could be observed at a point where *S* was poorly practiced and moderately well practiced. Therefore, the instructional variable was introduced at the first session. Since this procedure did not allow for matching the three groups, *S*s were assigned to groups by use of a table of random numbers.

Instructions.—The three groups differed in instructions given concerning the making of overt errors. For the high-error group it was stressed that *S* should try to make a response every time an item appeared. Members of this group were told that errors would not hurt their scores and that responses should be made even if they were only guesses.

In contrast, the low-error group was told that they should not respond unless they were certain that the response was correct. This group was told to say nothing rather than guess.

The control group was given instructions which neither told *S* that he should make errors nor not make errors. In effect, these instructions allowed *S* to respond "naturally."

In order to emphasize and maintain these instructions, abbreviated portions were repeated after every fifth trial during the learning of all three lists. The drum was stopped for 10 sec. while such instructions were given.

Materials.—The lists were composed of 14 nonsense syllables of low intralist similarity. These lists have been reproduced in an earlier publication (4). The first syllable was used as an anticipatory cue so that only 13 syllables were actually learned. The lists were presented on a Hull-type drum at a 2-sec. rate, with a 2-sec. break between trials except on every fifth trial (as noted above) where the interval was 10 sec. The learning of each list was carried to one errorless trial.

Since there were three independent groups of *S*s, each group learned the same three lists. The lists within groups were completely counterbalanced so that differences in difficulty were equally distributed over the three stages of practice.

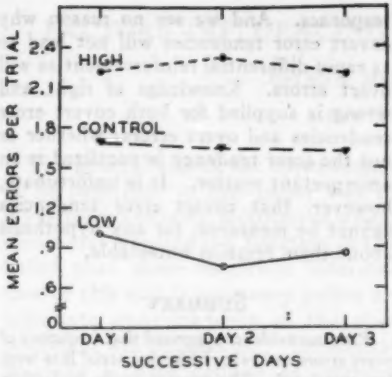

FIG. 1. Overt error frequency as a function of instructions concerning the making of errors

RESULTS

Overt errors.—We ask first whether or not the differential instructions concerning the making of overt errors are reflected in the frequency of errors. The answer to this question is given in Fig. 1. In this figure, stage of practice refers to the three lists learned on successive days. The ordinate measure was determined by dividing the total overt errors made in learning a given list by the number of trials required to learn the list for each *S*. The mean of these 24 values was used in plotting a single point in Fig. 1. As Fig. 1 indicates, large and consistent differences occurred in overt error frequency. Variance among distributions was very heterogeneous. As crude statistical tests of differences, Edward's approximation method (2), and also the median test as suggested by Moses (3), were applied to Day 1 scores. Both methods showed the differences to be highly significant. This evidence, plus the apparent correlation between the nature of the instructions and error frequency, leads to the conclusion that differences between groups in Fig. 1 are probably reliable.

Mean trials to learn.—Having shown that the three groups differed consistently in overt error frequency per trial, we ask the next question, which concerns the rates of learning of the three groups. The mean number of trials to learn for each group for each stage is shown in Fig. 2. Although the figure suggests that the control group, with "natural" error frequency, learns more rapidly than either of the other groups, especially in the initial stages, the differences are not reliable. Table 1 shows that no variable except practice reaches an acceptable level of significance. The method of analysis in Table 1 is that proposed by Archer (1). For evaluating uncorrelated observations *Ss*/Sequence/Instructions is the appropriate term; the pooled-interaction term is used to evaluate the correlated measures. No *F* has been entered which is less than unity. We must conclude that although instructions produced differences in overt error frequency per trial, these instructions had no appreciable influence on rate of learning.

We have also made analyses of error frequencies and acquisition rates at various stages of learning each list. None of these analyses showed any

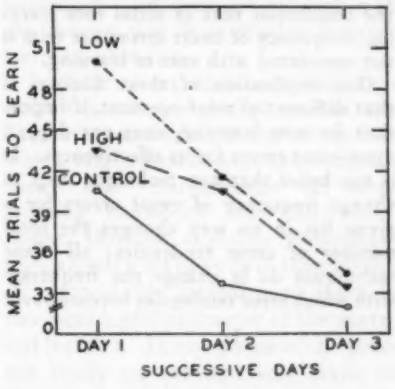

FIG. 2. Rate of learning as related to instructions concerning the making of errors

TABLE 1

ANALYSIS OF VARIANCE OF TRIALS TO LEARN

Source	*df*	Mean Square	*F*
Instructions	2	623.39	
Sequence	5	137.12	
Instr. × Seq.	10	512.53	
Ss/Seq./Instr.	54	710.98	
Practice	2	2478.43	31.62
Prac. × Instr.	4	105.99	1.35
Prac. × Seq.	10	139.22	1.78
Pooled interaction	128	78.37	
Total	215		

differences among conditions not demonstrated in Fig. 1 and 2.

DISCUSSION

The present results show that there is no significant association between rate of learning serial-nonsense lists and number of overt errors per trial made in learning those lists. In the present study wide differences in error frequency were produced by the different instructions to *S*. In a previous study (4), differences in error frequency resulting from different activities introduced between trials were also shown to be unrelated to rate of learning. Finally, correlations between rate of learning and frequency of overt errors per trial were about zero in this prior study. All these findings lead to the conclusion that in serial rote learning, frequency of overt errors per trial is not associated with rate of learning.

One implication of these findings is that differential reinforcement, if important for rote learning, does not depend upon overt errors for its effectiveness. It is our belief that any technique used to change frequency of overt errors for a given list in no way changes the total number of error tendencies; all these techniques do is change the frequency with which error tendencies become overt responses. And we see no reason why covert error tendencies will not lead to as rapid differential reinforcement as will overt errors. Knowledge of right and wrong is supplied for both covert error tendencies and overt errors; whether or not the error tendency is vocalized is an unimportant matter. It is unfortunate, however, that covert error tendencies cannot be measured, for any hypothesis about them remains untestable.

SUMMARY

Previous evidence suggested that frequency of overt errors and rate of learning serial lists were not associated. The present study undertook to produce wide differences in overt error frequency by appropriate instructions concerning the making of errors. One group was instructed to respond as often as possible by guessing; another group was instructed never to say a response unless they were sure it was correct. A third group was allowed to respond normally. Each group of 24 *Ss* learned three serial lists of nonsense syllables, one each day, for three days.

The results show that marked differences in overt error frequency per trial were produced by the instructions but that no measurable differences in rate of learning were associated with error frequency per trial. It is concluded that differential reinforcement in serial learning does not depend upon frequency of overt errors and that such reinforcement will take place as effectively when the errors are covert.

REFERENCES

1. ARCHER, E. J. Some greco-latin analysis of variance designs for learning studies. *Psychol. Bull.*, 1952, 49, 521–537.
2. EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1951.
3. MOSES, L. E. Non-parametric statistics for psychological research. *Psychol. Bull.*, 1952, 49, 122–143.
4. UNDERWOOD, B. J. Studies of distributed practice: VI. The influence of rest-interval activity in serial learning. *J. exp. Psychol.*, 1952, 43, 329–340.

(Received June 18, 1953)

Journal of Experimental Psychology
Vol. 47, No. 3, 1954

FACTORS INFLUENCING INCIDENTAL LEARNING[1,2]

GEORGE HASKELL BROWN[3]

New York University

The general aim of the experiments to be described in this paper was to investigate various conditions of which incidental learning is a function in the belief that more empirical information of this sort is necessary before an adequate understanding of the phenomenon will be possible. Incidental learning is here defined as learning which occurs in the *absence* of: (*a*) experimentally administered instructions to learn, and (*b*) introspectively reported intention to learn. An additional aim was to investigate the effect upon later motivated learning of an initial period of incidental learning, or, stated differently, the effect upon learning of the delayed introduction of instructions to learn.

The phenomenon of incidental learning has been of theoretical interest chiefly because it has often been regarded as an instance of unmotivated learning. Determining whether this is or is not the case is beset with many difficulties. Although it is relatively easy to demonstrate that learning (usually in small amounts) can occur without reportable intention to learn, it can always be argued that certain unconscious motives existed, or that some unknown secondary reinforcement agents were involved. McGeoch (6), for example, has suggested that whenever a set-to-read or a set-to-observe is aroused by instructions, it is to be expected that a set-to-*learn* will be evoked at the same time since throughout our past lives we have usually been set to learn that which we have read or observed and this learning has been rewarded.

In terms of Hullian learning theory, this notion involves an application of the concepts of stimulus generalization and secondary reinforcement. A situation similar to any situation in which motivated learning has occurred in the past may evoke the set to learn through stimulus generalization. The learning activities which are thus aroused may be in some measure maintained and rewarded by secondary reinforcement, i.e., by the appearance of stimuli which frequently have been experienced in association with rewarded past learning. Hence, a situation might be expected to possess stimulus generalization potency and secondary reinforcement properties to the degree that it is similar to past learning situations.

In order to test this hypothesis one must tentatively specify some of the characteristics of learning situations which might function in these capacities. One characteristic of many learning situations and one which might, through stimulus generalization, arouse a set to learn in the absence of a reportable intention to learn is the meaningful character of the material learned. In one phase of the present study an attempt was made to investigate this possibility by comparing the amounts of incidental learn-

[1] This paper is based on a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree in the Department of Psychology, Graduate School of Arts and Science, New York University. The writer is indebted to Professor L. W. Crafts for his generous advice and assistance throughout this investigation.

[2] Portions of this paper were presented at the 1953 meeting of the Eastern Psychological Association.

[3] Now at Mount Holyoke College.

ing shown by two groups, one of which learned three-letter English words, and the other of which learned three-letter nonsense syllables. An inferior score by the syllable group would support the stimulus generalization aspect of the theory considered above.

Another characteristic of many learning situations and one which might become a secondary reinforcer of verbal learning is the proprioceptive stimulation arising from overt pronunciation. If a group of *Ss* who silently observe a list of verbal items while making no effort to learn should show less incidental learning than a group which does pronounce the items aloud also while making no effort to learn, then overt pronunciation would appear to provide secondary reinforcement and the above theory would receive support.

Quite aside from these particular theoretical considerations, a worthwhile approach to the problem of incidental learning is to determine whether the so-called "laws of learning," which are derived from studies of motivated learning, also hold for incidental learning. If the two types of learning are fundamentally alike, then the conditions which affect intentional learning in characteristic ways should affect incidental learning similarly. If they should be found not to do so, then this fact in itself is important and essential to an adequate understanding of the phenomenon.

In our first and main experiment we compared incidental with intentional learning with regard to the influence upon them of (*a*) number of presentations of material, and (*b*) meaningfulness of material. An additional problem investigated was the effect upon the learning curve of the delayed introduction of instructions to learn.

Method

Apparatus.—A Hull-type memory drum was employed, using a 2-sec. exposure rate and an intertrial interval of 10 sec.

Learning materials.—The list of 12 nonsense syllables was composed of items having 0–20% association value according to Glaze's (3) determinations. The list of 12 English words was made up of monosyllabic nouns. Both lists were assembled in such a way that they would not lend themselves readily to the use of mnemonic devices.

Experimental design.—A three-variable factorial design with two variations of each variable was used in this experiment. The dependent variable was the number of correct anticipations on certain specified trials. The independent variables were (*a*) nature of instructions (incidental vs. intentional learning), (*b*) nature of material (words vs. nonsense syllables), and (*c*) number of presentations (four vs. eight).

Subjects and procedure.—The *Ss* were 160 undergraduate college students who were randomly assigned to eight groups of 20 *Ss* each. Each group was classified with respect to each of the three independent variables and will be identified by a three-item designation such as EN8 or CW4. The "E" or "C" indicates whether the group was experimental (incidental learning) or control (intentional learning). The "N" or "W" indicates whether the material was nonsense syllables or words. The "4" or "8" indicates the number of trials given before the learning test was introduced.

All *Ss* served individually. Each *S* of the experimental groups was initially informed that this was an experiment investigating the effects of fatigue upon speech and that he was to pronounce each of the serially presented items so that his speech could be recorded and studied. A "live" microphone and recording equipment were conspicuously present to make the situation more realistic. After a certain number of "trials" *S* was stopped, asked a few questions to determine whether he had decided on his own to start memorizing, and then instructed to start learning by the anticipation method to a criterion of two errorless trials. The total time consumed in conducting this inquiry and in explaining the anticipation procedure ranged between 2.75 min. and 3.25 min. The *S*'s first anticipation score provided a measure of incidental learning. His anticipation scores on subsequent trials would, of course, enable us to determine the effect upon the learning curve of the delayed introduction of instructions to learn.

Corresponding to each of the four experimental groups was a control group which was

TABLE 1

MEAN ANTICIPATION SCORES ON FIFTH AND NINTH TRIALS, RESPECTIVELY, FOR THE FOUR- AND EIGHT-TRIAL GROUPS

Material	Experimental Groups				Control Groups			
	4-Trial Groups		8-Trial Groups		4-Trial Groups		8-Trial Groups	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Syllables	.80	.69	1.55	1.10	2.30	1.26	2.90	1.74
Words	1.05	.89	1.50	1.36	4.70	1.87	6.00	1.94

told to start learning by the anticipation method from the start. In an effort to hold conditions as constant as possible for all groups, these control *S*s were interrupted after either four or eight trials (corresponding to the points at which experimental *S*s received instructions to learn) and administered a brief questionnaire. The results of the questionnaire were irrelevant to the purpose of the study.

RESULTS AND DISCUSSION

Mean anticipation scores for each of the groups are presented in Table 1. Before carrying out an analysis of variance the data were subjected to Bartlett's (2) test for homogeneity of variance. The result, $\chi^2 = 27$, $p < .01$, indicated that the variances were not homogeneous and, hence, that our data did not fulfill one of the conditions for the use of this method of analysis. No transformation proved satisfactory which, according to Edwards (2), is to be expected when some of the means are less than 2.0 as was the case in this study. Lindquist (4) suggests that even in the presence of heterogeneity of variance it may be permissible to use the analysis of variance technique provided that one insists on very highly significant *F*'s before generalizing from the data.

TABLE 2

ANALYSIS OF VARIANCE OF MEAN ANTICIPATION SCORES

Source	*df*	Mean Square	*F*
Instructions (A)	1	302.50	142.28*
Material (B)	1	81.22	39.81*
Trials (C)	1	24.02	11.77*
A × B	1	70.22	34.42*
A × C	1	1.22	
B × C	1	.40	
A × B × C	1	2.52	1.23
Within groups	152	2.04	

* Significant beyond the .001 level.

The results of the analysis of variance of the mean anticipation scores are presented in Table 2. All but one of the calculated *F*'s are significant beyond the .001 point. With 1 and 152 *df*, an *F* of 11.39 is needed for significance at the .001 level. In view of the magnitude of these *F* ratios it was decided to reject the null hypothesis and compute *t* tests between the groups. In doing so the standard error of the difference was based upon the variances of the two distributions being compared rather than upon the variance within groups indicated in Table 2. This is the procedure recommended by Lindquist (5) when heterogeneity of variance exists.

Effect of instructions.—The highly significant *F* for instructions means that intentional learning was vastly superior to incidental learning. Examination of Table 1 shows that each of the control group means is superior to its corresponding experimental group mean. The one-tail *t* tests for these differences were all significant beyond

the .01 point. For EW4 vs. CW4, $t = 7.93$; for EW8 vs. CW8, $t = 8.49$; for EN4 vs. CN4, $t = 4.69$; for EN8 vs. CN8, $t = 2.97$.

The superiority of intentional to incidental learning is not surprising since virtually all previous studies have found that to be the case. One might suppose that in this particular experiment it was due partly to the fact that the experimental *Ss*, at the time of their first test, were handicapped by their unfamiliarity with the anticipation technique. Partly in order to evaluate this possibility, an additional control group (CW4a) which would also suffer this handicap was run. These *Ss* were told to try to learn the same list of serially presented words that had been used with group CW4. Although they were told to expect a test, nothing was said to them regarding the nature of the test to be given. After four trials they were instructed on the anticipation technique which they were to utilize thereafter. This group's mean score of 3.00, $SD = 1.94$, on their first test trial is significantly smaller ($t = 2.83$, $p < .01$) than the score of CW4 in the main experiment. This suggests that unfamiliarity with the anticipation technique was a factor involved in the relative inferiority of the incidental learning groups in the main experiment. However, if we compare the incidental learning score of the original EW4 group with the intentional learning score of this new control group, intentional learning is still seen to be significantly superior ($t = 4.06$, $p < .01$). Although this interpretation of the inferior performance of the incidental learning groups is strictly applicable only to group EW4, since the new control group corresponds most closely with it, it seems reasonable to suppose that a similar explanation is appropriate with the other three experimental groups.

Effect of overt pronunciation.—In order to further elucidate the nature of incidental learning it was decided to study the effect upon it of overt pronunciation. An additional experimental group of 20 *Ss* was set up which could not be incorporated into the factorial design of the main experiment. The *Ss* were given eight exposures to the list of syllables under the pretext of having them rate them for aesthetic value. They indicated whether they liked or disliked each item by pressing certain buttons, thus obviating the need for overt pronunciation. After eight trials each *S* was instructed on the anticipation technique to be used on the next trial. The mean score for this group was 1.26, which was not significantly different from the score of group EN8 in the main experiment ($t = .88$, $p > .20$). Thus, pronunciation was found not to be a factor in incidental learning in spite of the fact that several experiments cited by McGeoch (6, p. 169) showed that with intentional learning the presence of overt pronunciation facilitated learning. It appears, therefore, that incidental learning differs from intentional learning in this respect.

Effect of meaningfulness of material.—The highly significant *F* for material indicates that words were more easily learned than syllables. However, the significant *F* for interaction between instructions and material indicates that the effect of meaningfulness differed for the two types of instruction, or, what amounts to the same thing, that the effect of instructions differed for the two materials. Further analysis of the data showed that for intentional learning, both the four- and eight-trial groups that learned words did significantly better than the corresponding groups that

learned syllables. The t's were, respectively, 4.80 and 5.34, $p < .01$ for both. With incidental learning, however, the four- and eight-trial word groups did not differ significantly from the corresponding syllable groups. The t's were, respectively, .81 and .13, $p > .40$ for both.

It has been known since Ebbinghaus that words are more easily learned than syllables. But this well-established generalization (6) was derived from studies of intentional learning and may not be applicable, at least not in the same way, to incidental learning. Actually, it does seem likely that a significant relationship between amount of incidental learning and meaningfulness of material could be demonstrated if meaningfulness were varied over a wider range, e.g., by comparing the incidental learning of syllables with that of a "series" of words which together made up a complete sentence. The present results, therefore, are not construed as demonstrating the absence of any relationship between these two variables but rather that the hypothetical curve relating ease of learning to meaningfulness of material has a different form for incidental learning from its form for intentional learning. In the latter a small difference between two materials in their meaningfulness will be reflected in their difficulty for learning; in incidental learning it appears that the difference must be much greater to have such an effect.

Effect of number of presentations.—The highly significant F for trials indicates that eight trials generally produced more learning than four trials. Examination of Table 1 reveals that each of the eight-trial groups scored higher than its corresponding four-trial group. In some instances the t's were significant; in other instances they were not. However, the failure of some of the t's to reach significance is probably a matter of chance since the F's for interaction between instructions and trials, and between material and trials, did not approach significance (see Table 2). Thus, we have no reason to suppose that there is any difference between incidental and intentional learning with regard to the influence of number of presentations.

Effect of delayed introduction of instructions to learn.—This phrase may call to mind the now classic latent learning study of Blodgett (1) in which he demonstrated that rats who had been traversing a maze without food reward for a number of trials showed a sharp drop in their error scores immediately after the introduction of food reward. Blodgett considered that the size of this sudden improvement was too large to be reasonably attributed to one-trial learning, and hence regarded it as evidence of latent learning.

It is interesting to examine the data of this experiment on human verbal learning to see whether or not a sort of "latent learning spurt" is evident. The following criterion of latent learning was initially adopted: Is the improvement of the experimental group immediately after the introduction of instructions to learn significantly larger than the improvement of the corresponding control group between the first and second trials? We chose the gain between the first and second trials as a basis for comparison because we felt that this pair of trials would be most likely on the average to produce greatest gains.

It should be noted that the experimental group gains with which we shall be concerned are those which occur between the first two trials following the introduction of instructions to learn. By the nature of the experiment we could have no measure of performance prior to the time S was told to start anticipating. This improvement may not be strictly comparable to that considered by Blodgett since he compared the scores for his animals on the last trial before food reward with the first trial afterward. In

TABLE 3

IMPROVEMENT SCORES FOR EXPERIMENTAL GROUPS FOLLOWING FIRST INSTRUCTED TRIAL AND FOR CONTROL GROUPS BETWEEN INDICATED TRIALS

Material	Experimental Groups				Control Groups							
	4-Trial Groups		8-Trial Groups		Trials 1 and 2				Trials 2 and 3			
					4-Trial Groups		8-Trial Groups		4-Trial Groups		8-Trial Groups	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Syllables	1.30	.98	.70	1.08	.60	.67	.30	.47	.65	.93	.65	.74
Words	2.05	1.19	1.75	1.52	2.00	1.12	1.95	1.23	.85	1.46	.90	1.37

any case, our experiment is analogous to Blodgett's in the sense that both are concerned with the effect upon performance of an increase in the level of motivation.

As to the evidence for "latent learning," Table 3 shows that, following their first instructed trial, three of the four experimental groups registered gains which were larger than the control group gains specified in the criterion. However, an analysis of variance yielded a nonsignificant F for instructions, thus indicating that these differences were not significant.

Perhaps the question should be regarded as settled at this juncture, in view of these negative findings. However, it occurred to the writer that the criterion employed was actually quite a rigorous one and that a more lenient one might yield results of interest. Consequently, a second criterion was set up which consisted of comparing the experimental group gains with the gains made by the control groups between the *second* and *third* trials. The appropriate data for these comparisons also are presented in Table 3. The over-all significance of these differences was evaluated by an analysis of variance. Despite the fact that Bartlett's (2) test for homogeneity of variance yielded a corrected χ^2 of only 14.88 (which is barely significant at the .05 level), it was decided to carry out the analysis, accepting as significant only those F's having p values beyond the .001 point.

Our main purpose in carrying out this analysis was to determine whether the experimental group gains significantly exceeded the control group gains. Table 4 shows that the F for instructions is highly significant ($p < .001$). Specific comparisons by means of t tests showed that three of the differences were significant. (EN4 vs. CN4, $t = 2.05$, $p < .05$; EN8 vs. CN8, $t = .17$, $p > .05$; EW4 vs. CW4, $t = 2.85$, $p < .01$; EW8 vs. CW8, $t = 2.73$, $p < .01$.) As can be seen in Table 4, none of the other F's was sig-

TABLE 4

ANALYSIS OF VARIANCE OF IMPROVEMENT SCORES

Source	*df*	Mean Square	*F*
Instructions (A)	1	18.91	13.41**
Material (B)	1	12.65	8.97*
Trials (C)	1	1.80	
A × B	1	1.80	
A × C	1	2.25	1.59
B × C	1	.30	
A × B × C	1	2.94	2.08
Within groups	152	1.41	

* Significant beyond the .01 level.
** Significant beyond the .001 level.

nificant according to the standards decided upon.

One might suppose that these large improvement scores here under consideration resulted from a rapid familiarization with the anticipation technique rather than from the change in instructions. This possibility can be evaluated by referring to data from the special control group (CW4a) which has been described earlier. This group did show a large gain following the first instructed trial ($M = 1.50$, $SD = 1.07$), but it was still smaller ($p < .07$) than that of the corresponding experimental group. Hence, it appears that the learning spurts considered above result, in some fashion, from the change in instructions. Thus, we may have something here which is akin to the Blodgett type of latent learning.

In a recent article Meehl and MacCorquodale (7) have attempted to incorporate the Blodgett effect within Hullian theory by supposing that the first food reward obtained by his rats served to increase their motivation so that the augmented D (drive), reacting multiplicatively with the small differential habit strengths already acquired, served to increase excitatory potential and hence performance. In our experiment we might reason similarly that the introduction of instructions to learn increased our *Ss*' motivation, and thus increased their excitatory potential, which led to improved performance on the next two trials.

Summary

Four experimental groups of 20 *Ss* each were given either four or eight trials of incidental learning of either words or syllables, the *Ss* pronouncing the items for the alleged purpose of having their speech recorded and studied. An additional experimental group was treated comparably except that no pronunciation was involved. Immediately following the incidental learning trials, *Ss* were required to begin learning by the anticipation method to a criterion of two errorless trials. Their first anticipation scores were taken as measures of incidental learning. Corresponding to each experimental group was a control group of 20 *Ss* that learned with intention from the start. The principal results were as follows:

1. Intentional learning was consistently superior to incidental learning under all conditions of this experiment.

2. Overt pronunciation was found not to be a factor influencing incidental learning. This contrasts with previous studies of *intentional* learning where pronunciation has been found to be a facilitating factor.

3. Words were no easier to learn than syllables under the conditions of incidental learning, whereas with intentional learning words were the easier material.

4. Significantly more learning was accomplished by eight trials than by four trials, with both intentional and incidental learning.

5. A phenomenon akin to the Blodgett type of latent learning was demonstrated with human *Ss* in a verbal-learning task by the delayed introduction of instructions to learn.

References

1. Blodgett, H. C. The effect of the introduction of reward upon the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1929, **4**, 113–134.
2. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
3. Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.
4. Lindquist, E. F. *Statistical analysis in educational research.* Boston: Houghton Mifflin, 1940.
5. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
6. McGeoch, J. A. *The psychology of human learning.* New York: Longmans, Green, 1942.
7. Meehl, P. E., & MacCorquodale, K. Drive conditioning as a factor in latent learning. *J. exp. Psychol.*, 1953, **45**, 20–24.

(Received June 12, 1953)

Journal of Experimental Psychology
Vol. 47, No. 3, 1954

INCIDENTAL LEARNING UNDER TWO INCENTIVE CONDITIONS

HARRY P. BAHRICK

Ohio Wesleyan University[1]

Incidental learning has been defined by McGeoch and Irion as "learning which apparently takes place without a specific motive or a specific formal instruction and set to learn the activity or material in question" (2, p. 220). The word *apparently* has been included in this definition because experimental findings (1, 3, 4) indicate that when such learning occurs it is frequently the result of self-instructed sets which are brought to the situation by *S*.

The great difficulty of preventing motives not induced by *E* from influencing the behavior of *S*s has been demonstrated most clearly in the latent-learning controversy. Many investigators have attempted to perform crucial experiments to answer the question whether learning can occur in the absence of need reduction, and yet this question remains unanswered. It is apparently not possible to specify what needs are being reduced in a situation even when rats are the *S*s of the experiment. The relation between motivation and learning will nevertheless continue to be investigated because it is fundamental to the development of behavior theory. It would seem perhaps that a more fruitful approach to this problem consists of testing functional relations between relevant and irrelevant learning and those variables controlling motivation which *E* is capable of manipulating.

[1] The writer wishes to acknowledge many helpful suggestions by Dr. David Bakan, Dr. Paul Fitts, and Dr. Merrill Noble, and the help of Miss Dorothy Alexander, Miss Eleanor Bryenton, and Miss Carolyn Hall in the collection of data.

The present experiment attempts to investigate the relation between the amount of incidental learning and the strength of the *E*-induced motive in a learning situation. Specifically, the hypothesis being tested is that the amount of learning irrelevant to an *E*-induced set is inversely related to the strength of the incentive used by *E* to determine that set.

Using the terminology of sign-gestalt learning theory the following rationale would lead to the above hypothesis: If a set is looked upon as a response process, it is determined, like other acts of behavior, by goal demands and cognitions relating the act to the attainment of goals. If *S*s are simultaneously under the influence of several incompatible goal demands one would expect frequent changes in behavior, provided that the goal demands were sufficiently equal in strength to permit first one and then another to dominate. Of course, inhibition, fatigue, and other factors would also play an important role in determining variability of sets. Only in the case where one of the coexisting goal demands is much stronger than the others, and where relatively firm cognitions with respect to the satisfaction of that goal demand already exist, would one expect the persistence of a particular kind of set, or overt act, at the expense of other behavior relevant to the satisfaction of the other coexisting goal demands.

The above considerations lead to the experimental hypothesis. Thus, we would expect that learning irrele-

vant to an *E*-induced set will take place whenever the goal demand determining that set is temporarily exceeded by the strength of some *S*-induced goal demand determining another set. This is least likely to happen when the *E*-induced goal demand is particularly strong, as is the case when the incentive determining that goal demand is increased.

Method

Subjects.—Seventy-four *S*s were selected from elementary psychology classes and were assigned alternately to a high incentive group and a low incentive group of 37 *S*s each.

Apparatus.—The apparatus was a memory drum which serially exposed 14 geometric forms at a 3-sec. rate of presentation. Seven different forms were each repeated twice, and each form was filled with one of seven colors to make the following series of 14 items: red triangle, gray moon, blue cross, green circle, red moon, black rectangle, blue square, yellow circle, green diamond, brown rectangle, yellow triangle, black square, gray diamond, brown cross. Each figure was drawn so as to occupy an area of approximately ¼ sq. in.

Experimental procedure.—Both groups were given the same instructions which described the task of anticipating each form, beginning with the second trial, before the form appeared in the drum window. Colors were not mentioned in the instructions. Additional instructions intended to produce differential incentive to learn were then given.

The high incentive group was told: "We want to see how well you can do this task when you are trying your very best. We have been authorized to pay you a bonus ranging from 10¢ to $1.50, the exact amount depending upon how many times the list has to be presented before you can anticipate every one of the 14 forms correctly."

The low incentive group was told: "We are interested in finding out whether this task can be learned when you are not trying very hard. Try to pay just enough attention to keep up with the task, as you would in listening to a lecture which does not interest you."

Training continued for both groups until a criterion of one perfect trial was reached. Fifteen-second rest intervals were given between trials.

Immediately after *S*s had reached the criterion they were shown a 14-row × 7-column chart showing the forms along the rows in the same sequence as during learning, and each of the seven colors along the columns. They were instructed to point out the color associated with each form.

TABLE 1

Means for Form and Color Learning

Incentive	Trials to Learn Forms (T)		Correct Color Recognitions (C)		Adjusted Means for Color Recognition (C')
	Mean	*SD*	Mean	*SD*	Mean
High	16.9	4.21	6.14	2.56	6.17
Low	19.6	4.98	7.62	2.29	7.59

Results and Discussion

Table 1 shows the means and standard deviations of the two groups on the number of trials to reach the criterion (T), the number of correct color recognitions (C), and the adjusted number of correct color recognitions (C′), where the adjustment was made on the basis of the regression of C on T.

Table 2 presents the necessary information for the tests of significance. The difference between the means of

TABLE 2

Analysis of Covariance of Color Recognition Data Controlling Number of Trials to Reach the Criterion of Form Learning

Source	Unadjusted			Adjusted	
	df	Mean Square$_T$*	Mean Square$_C$†	*df′*	Mean Square$_{C'}$‡
Between groups	1	132.45	40.89	1	34.35
Within groups	72	21.84	6.07	71	6.14
Total	73			72	

* Mean Square$_T$ = Mean square for trials to reach the criterion of form learning.
† Mean Square$_C$ = Mean square for color recognitions.
‡ Mean Square $_{C'}$ = Adjusted mean square for color recognitions.

the two groups on T is significant at the 2% level of confidence ($F = 6.06$). As expected, the rate of learning relevant to the *E*-induced set was higher for the group offered an incentive.

The means of *both* groups are far above chance on C, with two correct responses out of 14 being expected on the basis of chance alone. More important for our hypothesis is that the C means differ from each other significantly at the 2% level of confidence ($F = 6.74$).

Since the low incentive group received on the average a larger number of trials to reach the criterion, it is possible that the greater amount of incidental learning of that group is the result of these extra trials and not of the low incentive offered during training. The analysis of covariance was done to control for this possible effect of T on C, especially since the difference between the groups on T was found to be significant. However, even after the adjustment is made the difference between the means is still significant at the 2% level of confidence ($F = 5.59$), indicating that even though there was a significant difference between the groups on T, it had no appreciable effect on the results of color recognition. This is best explained by the low correlations between T and C. The correlations between T and C were .048 and .126 for within-groups and total, respectively.

Summary

This experiment tests the hypothesis that the amount of learning irrelevant to an *E*-induced set is inversely related to the strength of the incentive determining that set.

Two groups of 37 *S*s learned a sequence of 14 colored geometric forms to one perfect trial by the method of serial anticipation. Their instructions were to learn the names of the forms only. At the end of this training they were tested for incidental learning of the colors used in the forms. The high incentive group was offered a money bonus depending upon the speed of learning; the low incentive group was told not to try very hard.

Results show that the learning of the form sequence was faster under the high incentive condition, but that the high incentive group scored significantly lower on the color recognition test. The difference was still significant ($p = .02$) when the variable of rate of learning of the form sequence was controlled by analysis of covariance. This was interpreted to support the hypothesis that led to the experiment.

References

1. Bahrick, H. P., Rankin, R. E., & Fitts, P. M. Effect of incentives upon reactions to peripheral stimuli. *J. exp. Psychol.*, 1952, **44**, 400–406.
2. McGeogh, J. A., & Irion, A. L. *The psychology of human learning.* New York: Longmans, Green, 1952.
3. Postman, L., & Senders, V. L. Incidental learning and generality of set. *J. exp. Psychol.*, 1946, **36**, 153–165.
4. Shellow, S. M. Individual differences in incidental memory. *Arch. Psychol., N. Y.*, 1923, **10**, No. 64.
5. Tolman, E. C. There is more than one kind of learning. *Psychol. Rev.*, 1949, **56**, 145–155.

(Received June 23, 1953)

Journal of Experimental Psychology
Vol. 47, No. 3, 1954

LEARNING RESPONSE COMPOUNDS HAVING TWO CRITICAL COMPONENTS

ALVIN J. NORTH AND JAMES HARRINGTON, JR.
Southern Methodist University

This study was concerned with the learning of a set of response compounds having two critical components. A response compound is a response which has two or more components, parts, aspects, dimensions, etc. For example, an arm movement on a reaction panel might have distance and direction components. A sorting task might call for placing objects in the proper row and column. A control knob might have to be rotated in a certain direction and by the right amount. The concept of response compounds is implied in work on dimensional analysis of motion (4) and response codes (1).

The concept of *critical* components in response compounds refers to those components whose values in *combination* determine rightness or wrongness of response. For example, two components are critical if each unique combination of their values is to be associated with a different signal.

The response compounds of this study varied in two aspects, namely, distance and direction from a central point on a reaction panel (see Fig. 1). There were three directions and three distances of movement, and each of the nine direction-distance combinations was to be associated with a different stimulus. Note that each response was at the same distance as two others and also in the same direction as two others. Subsequent theoretical analysis will show that this fact complicates the learning process.

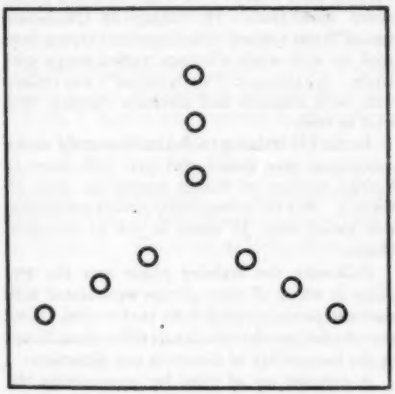

FIG. 1. Diagram of the reaction panel. Coding directions by "U" for up, "R" for right, and "L" for left, and coding distances from the center by "1," "2," and "3," the stimuli and their responses were as follows: Sam-U1; Joe-U2; Bob-U3; Walt-R1; Ed-R2; Carl-R3; Don-L1; Herb-L2; Nick-L3.

PROBLEM

How is such a set of response compounds having two critical components learned? Following stimulus-response theory, two assumptions are made: (*a*) discrimination learning involves differentiation of habit strengths as a result of differential reinforcement and nonreinforcement, and (*b*) the several components of a response compound acquire habit strength independently.

A response correct in both direction and distance is always reinforced, but a response incorrect in either or both components is nonreinforced. The task is complicated, however, by the fact that the correct direction is reinforced on some trials and nonreinforced on others, depending on the distance with which it is combined. Likewise, the correct distance depends on the concurrent direction for its reinforcement or nonreinforcement. Hence each component of the correct response compound is both aug-

mented and weakened in habit strength. Response compounds incorrect in *both* distance and direction would involve no difficulties beyond the usual generalization of excitatory and inhibitory tendencies along response dimensions.

Now consider the effects of two kinds of training: (*a*) "equal variation," in which both components are varied from trial to trial, and (*b*) "unequal variation," in which one component is varied from trial to trial while the other is varied from block to block of trials.

Consider equal variation first. Suppose distances and directions were equally learnable. With distribution of responses, errors in distance and direction would be about equally frequent. Consequently, the two components would have similar distributions of reinforcements and nonreinforcements. Hence the habit strengths of the several response components would differentiate at about equal rates. Therefore directions would be learned as accurately as distances.

Turning next to training with unequal variation, suppose distance was varied from trial to trial while direction was varied only from block to block. Learning something of the probability that correct responses tend to be in a given direction until a new direction is indicated, most of *S*'s responses would be in the correct direction. Therefore most of his errors would be in distance (direction being correct). Because of more frequent nonreinforcement, incorrect distances would become more strongly inhibited than wrong directions. Following such training, if *both* distance and direction were varied "randomly," responses would be more accurate in distance than in direction. If, on the other hand, direction were varied more frequently in training than distance, directions would be learned more accurately than would distances. Note that the component having more errors in training would become better learned than the component in which *S* usually is correct.

From the above theoretical analysis derives the main hypothesis of the study: Given equal learnability of two response components, the component which is varied more frequently in training becomes more accurately learned.

Method

The set of response compounds selected for study consisted of movements varying in three directions and three distances from a central point on a reaction panel (see Fig. 1). Each movement was to be associated with a different stimulus. The learning of the two critical components was investigated as a function of their relative rates of variation.

Subjects.—The *S*s were 30 university student volunteers, 17 men and 13 women, varying in age from 17 to 33 yr., with a mean of 18.6 yr.

Apparatus.—The apparatus consisted of an 18-in. square reaction panel with nine jewel lights arranged as in Fig. 1. The positions were arranged along three directions (including equal central angles) and at three distances from the center (at 1½-in. intervals). Each jewel light could be lighted independently by *E* from a control panel.

Nine common monosyllabic boys' names, differing in initial letter, were used as stimuli. Each name stimulus designated a different distance-direction combination in movement.

Design.—The experiment consisted of a single session having two phases: (*a*) a training phase, consisting of 135 trials, and (*b*) a test phase, also having 135 trials. A 10-min. rest period was given between training and test phases. Each *S* was assigned at random to one of the following groups, each of which had 10 *S*s: (*a*) Group A ("distance varied") was trained with distance varying from trial to trial while direction varied every nine trials. (*b*) Group B ("direction varied") was trained with direction varying from trial to trial while distance varied every nine trials. (*c*) Group C ("both varied") was trained with both distance and direction varying from trial to trial.

In the 135 training trials the frequently varied component was varied well over 100 times (a limited number of chance repetitions were allowed). But the infrequently varied component was varied only 14 times in the 15 nine-trial blocks.

Following the training phase was the test phase in which all three groups were tested with both components varied from trial to trial. Note that the design counterbalances differences, if any, in the learnability of distances and directions.

A complex set of rules for constructing the sequences was devised. These rules were intended to equalize the distribution of eligible stimuli within nine-trial blocks, avoid periodicities and frequent repetitions, etc.

Procedure.—The *S* was seated at a table on which was the reaction panel (Fig. 1). The panel surface was an inclined plane which *S* viewed at an angle. He was told to rest his preferred hand (all *S*s were right handed) on the reaction panel with his index finger at the center. The panel

was oriented as a sheet of writing paper so as not to obscure any of the jewel lights. The *S* was instructed that the jewel lights represented the positions of nine boys in a seating arrangement. He was told that upon hearing a name he was to point to and touch the appropriate jewel light and then return his hand to the center. Guessing was encouraged. If the response was wrong or failed to occur within 4 sec. (measured by clicks from a metronome), *E* flashed the jewel light at the correct position. If *S* responded correctly, no confirming signal was given. The names were presented at 5-sec. intervals, having been recorded on wire, the order depending on the training condition. At the beginning of the test phase, *S* was told that no further information would be given about rightness or wrongness of responses.

The *E* recorded *S*'s responses. If, as infrequently happened, more than one response occurred on a given trial, only the final response was recorded. More complete recording of anticipatory responses, although desirable, was impracticable in view of *E*'s other tasks.

Results

Responses were classified into four independent types: *Type 1*, direction correct and distance correct; *Type 2*, direction correct and distance incorrect; *Type 3*, direction incorrect and distance correct; and *Type 4*, direction incorrect and distance incorrect. Various combinations of these types were utilized in analysis. Note that Types 1 and 2 are correct in direction; that Types 1 and 3 are correct in distance; that Types 2 and 4 are incorrect in distance; and that Types 3 and 4 are incorrect in direction. The percentage distribution of response types is shown in Table 1. Since most data distributions were skewed, tests of statistical significance assuming normality were not used.

The hypothesis to be tested specifies equal learnability of response components. The learnability of distances and directions may be compared by comparing the percentage of errors in distance with the corresponding percentage for direction using the data of Group C (both varied). In the training phase, 33.6% of the responses erred in distance (Types 2 and 4) and 35.2% in direction (Types 3 and 4). The corresponding data for the test phase are 5.2% errors in distance and 5.3% in direction. These differences are small and reversed in direction. Combining the data of the two phases, six *S*s had more distance errors and four had more direction errors. On the average, then, distances and directions were about equally learnable, so that the condition specified in the hypothesis was satisfied.

The theoretical analysis assumed that during training Groups A and B would have more errors in the frequently varied component. The data support this expectation: Group A had 13.7% distance (Types 2 and 4) errors, but only 6.5% direction (Types 3 and 4) errors. Group B had 14.3% direction, but only 7.4% distance, errors. During training all 20 *S*s in Groups A and B had more errors in

TABLE 1

Percentage Distribution of Response Types

Group	Training Phase				Test Phase			
	Response Type*				Response Type*			
	1 (++)	2 (+−)	3 (−+)	4 (−−)	1 (++)	2 (+−)	3 (−+)	4 (−−)
A. Distance varied in training	84.4	9.1	1.9	4.6	92.2	2.0	3.5	2.3
B. Direction varied in training	82.5	3.2	10.1	4.2	93.0	3.5	2.4	1.1
C. Both varied in training	53.1	11.7	13.3	21.9	92.0	2.7	2.8	2.5

* Type 1, direction and distance correct; Type 2, direction correct and distance incorrect; Type 3, direction incorrect and distance correct; Type 4, direction and distance incorrect.

the frequently varied component. Consequently, because of differential nonreinforcement, Group A would be expected to inhibit incorrect distances more than incorrect directions in the test phase. Conversely, Group B would be expected to inhibit incorrect directions more than incorrect distances. The test phase data confirm these expectations and the main hypothesis: Group A had a lower percentage (4.3%) of distance (Types 2 and 4) errors than of direction (5.8%) errors (Types 3 and 4). Conversely, Group B had a lower percentage of direction than of distance errors (3.5% versus 4.6%). Although small in magnitude, the differences were consistent. Combining Groups A and B, 16 *Ss'* differences were in the predicted direction, 2 were not, and 2 were ambiguous (no difference). On the null hypothesis, 9 of the 18 unambiguous differences would be in the predicted direction and 9 would not. Comparing observed with expected frequencies, chi square equaled 9.38 (corrected for continuity), which is significant beyond the 1% level of confidence. Hence, the component which was varied more frequently in training, whether distance *or* direction, became more accurately learned. It should be noted that the level of learning was fairly high (92% to 93% correct in the test phase). Presumably, less training would have resulted in larger differences, thus providing a more sensitive test of the hypothesis.

Another line of analysis supported the assumption that the several components of a response compound independently acquire habit strengths. Theoretically, the habit strength of the direction component can be held constant by selecting only responses incorrect in direction (Types 3 and 4), since neither incorrect direction had been reinforced with respect to the stimulus. Assuming that response components independently acquire habit strengths, the distance component of such "direction incorrect" responses had been reinforced on other trials if correct and combined with the correct direction, but incorrect distances were never reinforced. Hence, the habit strength of responses incorrect in direction and correct in distance (Type 3) should be greater than that of responses incorrect in direction and incorrect in distance (Type 4), even though neither of these types of responses had ever been reinforced. Test phase data were analyzed to test this prediction. On any given trial, there were two Type 3 possibilities and four Type 4 possibilities (plus two Type 2 and one Type 1). "Relative percentages" then were calculated as follows: Type 3 percentages were divided by two, and Type 4 percentages by four. The test phase data of Table 1, when converted to relative percentages, agree with the hypothesis in the case of each group. Within each group most differences were in the predicted direction. Of the 30 *Ss*, 21 differences were in the predicted direction; 6 were not; and 3 were ambiguous. A test of the null hypothesis (on the unambiguous differences) yielded a chi square of 7.26, which is significant beyond the 1% level of confidence. Hence, the several distances acquired habit strengths independent of the habit strength of the direction component.

Analogously, responses incorrect in distances and correct in direction (Type 2) ought to have greater habit strength than responses incorrect in distance and incorrect in direction (Type 4). The test phase data support this prediction, with most of the individual differences in each group in accord with the hypothesis. Of the 30 *Ss*, 22 differences were in the predicted direction; 4 were not; and 4 were ambiguous. In testing the null hypothesis (using the unambiguous instances), a chi-square value of 11.11 was obtained, which is significant beyond the 1% level of confidence.

Hence with respect to a given stimulus the several directions had differential habit strengths independent of the habit strength of the distance component.

The foregoing results show that a response incorrect in one component was more likely to be correct than incorrect in the other component. This evidence thus supports the assumption that the several response components independently acquire habit strength.

Other findings.—Although Group C had markedly fewer (53.1%) reinforced responses in training than did Group A (84.4%) or Group B (82.5%), all three groups did about equally well in the test phase, having correct 92.0%, 92.2%, and 93.0%, respectively. The higher percentage of errors in Group C's training had no adverse effect on final performance, since, theoretically, the values of each component were differentially reinforced. No prediction was made comparing the final performance of Groups A, B, and C, since, presumably, any predictions would depend on the degree of training involved.

Discussion

The results appear to confirm the main hypothesis of the study, namely, given equal learnability of two response components, the component which is varied more frequently in training becomes more accurately learned. There also appeared to be evidence for the assumption that the several response components acquire habit strengths independently. Could these results, however, be explained by an alternative theory concerning the learning of response compounds, namely, that response compounds as *undifferentiated wholes* acquire habit strength? According to this hypothesis, the response as a *unit*, not its several hypothetical components, acquires habit strength.

This "unit" hypothesis is in agreement with certain results: It is consistent with the finding that Group C's distance and direction errors were about equally frequent. It is also consistent with, in fact predicts, the finding that in the test phase fewer responses would err in the component which was more frequently varied in training. But the hypothesis fails to predict the finding that in the test phase a response incorrect in one component was more likely to be correct than incorrect in the other component. Furthermore, this hypothesis logically implies the following propositions regarding performance in the test phase, which, however, are directly contrary to the results:

a. In Group A, responses erring in distance ought to be *less* probable in the correct than in the incorrect direction.

b. In Group B, responses erring in direction ought to be *less* probable in the correct than in the incorrect distance.

c. In Group A, responses erring in direction ought to be as probable in the correct as in the incorrect distance.

d. In Group B, responses erring in distance ought to be as probable in the correct as in the incorrect direction.

e. In Group C, responses erring in direction ought to be as probable in the correct as in the incorrect distance.

f. In Group C, responses erring in distance ought to be as probable in the correct as in the incorrect direction.

Proposition *a* will be derived for Group A, and its analogue, Proposition *b*, will then be regarded as proved for Group B. Group A was trained with distance varied more frequently than direction. During training, responses in the correct direction and incorrect distance (Type 2) occurred more frequently than responses incorrect in direction *and* distance (Type 4). Because of more frequent nonreinforcement, Type 2 responses would have weaker habit strengths than would Type 4 responses. Therefore, in the test phase, responses incorrect in distance (Types 2 and 4) would have *less* probability of being in the correct direction (Type 2) than in the incorrect direction (Type 4).

Proposition *c* will now be derived, and, as before, its analogue, Proposition *d*, will be considered proved. During training, Group A (distance varied) most frequently responded in the correct direction, and the relatively few errors in direction (Types 3 and 4) were distributed in distance. Being about equally often nonreinforced, Type 3 and 4 responses should have equal habit

strengths. Therefore, in the test phase, responses erring in direction (Types 3 and 4) ought to be as probable at the correct distance (Type 3) as at the incorrect distance (Type 4).

Proposition *e* will now be derived, and then its analogue, Proposition *f*, will be considered proved. During training Group C's errors in direction and/or distance were distributed. Hence, habit strengths of all erroneous responses, being about equally often nonreinforced, ought to be equal. Therefore, in the test phase, responses erring in direction (Types 3 and 4) ought to be as probable at the correct distance (Type 3) as at the incorrect distance (Type 4).

Since each of the six propositions is at variance with the obtained results, the hypothesis that response compounds acquire habit strengths as undifferentiated wholes, therefore, is rejected. Instead, the results support our assumption that response components independently acquire habit strength, acknowledging, of course, that some responses are more readily analyzable into components than are others. If the rejected hypothesis is modified to allow for "response generalization" by virtue of common distance or direction elements, then it becomes logically equivalent to the assumption of this study.

The results of the present experiment parallel those of North and Leedy (3), who studied learning to discriminate stimuli having two critical components. They found that the stimulus component varied more frequently in training becomes better discriminated. The concept that stimuli have components which independently acquire habit strengths with respect to responses is a familiar one. The present study, however, shows that *response components* independently acquire habit strengths with respect to stimuli. Although other studies have investigated the learning of responses depending on multiple stimulus contingencies (e.g., 2, 3), the study of multiple response contingencies, as in the present experiment, is relatively novel. Stimulus-response theory appears to be applicable to both stimulus and response contingencies. In further studies it would be desirable to examine the effects of other variables such as number of critical components, number of values per component, learnability or discriminability of components, frequency of variation, etc.

Summary

The learning of a set of responses having two critical components, distance and direction, was investigated. A theoretical analysis assuming equal learnability of distances and directions in the task predicted that the component varied more frequently in training would become better learned.

In the task, each of nine unique movements compounded of direction and distance components was to be associated with a different signal. Thirty *Ss*, divided into three equal groups, had 135 training trials. Group A was trained with distance varied more frequently than direction; Group B, with direction varied more frequently than distance; and Group C, with both components varied equally frequently. All groups were then tested for 135 trials with both components varied equally frequently.

The results show:

1. There was equal learnability of distances and directions in the task.
2. Given equal learnability of two response components, the component which was varied more frequently in training became more accurately learned.
3. A response that was incorrect in one component was more likely to be correct than incorrect in the other component. This evidence supported the assumption that the several response components acquire habit strengths independently.
4. Certain implications of the assumption that response compounds acquire habit strengths as undifferentiated wholes were shown to be inconsistent with the results obtained.

References

1. Fitts, P. M., & Seeger, C. M. S–R compatibility: spatial characteristics of stimulus and response codes. *J. exp. Psychol.*, 1953, **46**, 199–210.
2. Nissen, H. W. Analysis of a complex conditioned reaction in chimpanzees. *J. comp. physiol. Psychol.*, 1951, **44**, 9–16.
3. North, A. J., & Leedy, H. B. Discrimination of stimuli having two critical components when one component varies more frequently than the other. *J. exp. Psychol.*, 1952, **43**, 400–407.
4. Rubin, G., & Smith, K. U. Learning and integration of component movements in a pattern of movement. *J. exp. Psychol.*, 1952, **44**, 301–305.

(Received June 9, 1953)

Journal of Experimental Psychology
Vol. 47, No. 3, 1954

STUDIES OF THE EFFECT OF CHANGE OF DRIVE: III. AMOUNTS OF SWITCHING PRODUCED BY SHIFTING DRIVE FROM THIRST TO HUNGER AND FROM HUNGER TO THIRST

HOWARD H. KENDLER, ALAN D. KARASIK, AND ALAN M. SCHRIER[1]

New York University

It has been reported (3) that albino rats after learning to go to one end of a T maze to obtain dry food have a tendency to switch their spatial response when made thirsty, e.g., turn left if the turning habit acquired under hunger was to turn right.

In an attempt (5) to explain this switching phenomenon it was reasoned that *S*s in the original study had developed anticipatory eating responses which had become conditioned to the cues in the alley leading to food. It was assumed that (*a*) these fractional anticipatory eating responses conflicted with the resultant condition of water deprivation, and (*b*) that this incompatibility would be reduced by *S*'s going *away* from the alley leading to food. This explanation would follow because the cues in the incorrect path of the maze had not been directly conditioned to the anticipatory eating responses and consequently the greater the penetration into this path, the greater would be the relief from the conflict state postulated to exist.

The implication of the above theoretical analysis is that the tendency to switch is positively related to both the strength of the anticipatory eating response at the choice point and the intensity of the thirst drive. Diminishing the strength of either should diminish the tendency to switch; increasing the strength of either should increase the tendency to switch. One part of this prediction was confirmed in a study (5) which reported that *S*s required to achieve a criterion of 12 successive correct choices before being shifted to 22-hr. thirst drive evidenced 48% switching, while only 30% of the *S*s that reached a criterion of 3 successive correct responses switched. A second study (6) demonstrated that 45% of *S*s whose motivation was shifted to 22-hr. thirst drive switched, while only 21% of those *S*s switched whose motivation had been shifted to a 3-hr. drive.

The present experiment is concerned with testing a third implication of the avoidance hypothesis which has been offered to explain the switching phenomenon. It would be expected, according to this formulation, that more switching should occur when the motivational condition of the *S* is shifted from hunger to thirst, as compared to shifting from thirst to hunger. As will be discussed later, *S*s shifted from hunger to thirst would possess a tendency to avoid the side which they had previously chosen, whereas in those shifted from thirst to hunger no such avoidance tendency should exist.

[1] The first two studies presented in this paper were conducted by the two junior authors when they were students in the undergraduate experimental psychology course taught by the senior author. The third study was part of A. D. K.'s M.A. thesis, while both H. H. K. and A. D. K. are responsible for the article in its present form. A. M. S. is now a graduate student at the University of Wisconsin. The authors wish to express their appreciation to Dr. Tracy S. Kendler for her critical reading of the manuscript.

Procedure

The experimental procedures of the three replicated studies employed in this experiment are essentially the same as those used for the third and fourth studies of the initial experiment (5) and all three studies of the second experiment (6), except for the modifications required to test the above prediction. Since the general procedure is described fully elsewhere (5), only a brief description is presented here.

The total number of *S*s used in the three replications of the present experiment were 68 naive male albino rats, of the Wistar strain, purchased from the Albino Farms, Red Bank, New Jersey. Their ages ranged from 70 to 95 days at the beginning of the experiment.

The apparatus was a single-choice T maze. After preliminary training all *S*s were given three free-choice trials daily. Half of the *S*s were motivated by a 22-hr. thirst drive. The hungry *S*s had ten pellets of Purina dog chow in one end box and nothing in the other. The thirsty *S*s found a glass coaster filled with water while the other end box was kept empty. All *S*s were trained until they reached the criterion of four days of correct choices (at least 12 successive errorless trials). On the day following the last criterion trial, the motivation of *S*s was shifted; those *S*s who had been very thirsty were now hungry while those *S*s who had been hungry were now thirsty. This shifting of the drive state was accomplished for a hunger-to-thirst *S* by permitting *S* to drink for 30 min. following its last training trial and at the same time providing it with 5 gm. of food. After the water was removed, another 5 gm. of food was inserted into the cage. Thirty minutes before the test trial, a large handful of food was placed in the cage and *S* was permitted to satiate itself. The motivation of *S* in the thirst-to-hunger group was controlled by giving *S* 5 gm. of food following the last training trial, and at the same time allowing it to drink for 30 min. In order to insure satiation, water was made available to *S* for 1 hr. prior to the test trial. Each *S* of both groups received one test trial under the new drive state.

Results and Discussion

The *S*s who received their initial training while thirsty reached the criterion of learning more rapidly than did the hungry *S*s. The ranges of the mean number of errors in the three studies were from 2.0 to 3.3 for the thirsty *S*s and from 4.3 to 6.7 for the hungry *S*s. Equivalent figures for the mean number of reinforcements were 16.6 to 18.0 and 22.9 to 27.0 for the hungry and thirsty *S*s, respectively. It was found, with the use of the chi-square test of combining probabilities from independent tests of significance, that the differences between the two groups, in terms of both errors and reinforcements, were significant at the .01 level.

TABLE 1

Number and Percentage of *S*s Switching Their Spatial Response in Hunger-to-Thirst and Thirst-to-Hunger Groups

Study	Group H–T			Group T–H		
	N	Switching		*N*	Switching	
		N	%		*N*	%
1	7	4	57.1	8	2	25.0
2	9	3	33.3	9	2	22.2
3	18	7	38.9	17	1	5.9
Total	34	14	41.2	34	5	14.7

The main results of this experiment are contained in Table 1. They are consistent with the prediction made: that the hunger-to-thirst group (Group H–T) should demonstrate a greater percentage of switching than would the thirst-to-hunger *S*s (Group T–H). The differences, with the use of the one-tail hypothesis, were significant at the .10, .30, and .005 levels. A combination of these three probabilities by means of the chi-square test yielded an over-all significance beyond the .01 level. It would, therefore, seem justified to reject the one-tail null hypothesis that there is no difference in the amount of switching displayed by the two groups in favor of Group H–T over Group T–H.

An obvious question that might be raised pertains to the behavior of Group T–H on the test trial. Why

should this group demonstrate any switching behavior? The answer to this question lies in the analysis of the stimulus compound to which the turning response had been conditioned. Not only do the cues at the choice point become connected to the response, but the drive stimulus itself does also (2, 4, 7). When this drive stimulus is withdrawn, an important component of the stimulus compound previously conditioned to the correct turning response is removed and hence a drop from the previous 100% level of responding would be expected. The *S*s of Group T–H should persist in the execution of their learned response at an "above chance" level when the drive is shifted to hunger, since the cues at the choice point still remain and, because of the training series, are capable of evoking the learned turning habit.

In addition to the prediction regarding the relative amount of switching behavior of our two experimental groups, our theoretical analysis would lead us to predict that Group H–T should evidence, on the test trial, a greater disturbance in its running time than would Group T–H. This would follow since it is postulated that Group H–T is subjected to a conflict situation involving strong tendencies to enter both alleys of the T maze, while no such similar conflict between two opposing responses would exist for Group T–H. We may use as a measure of disturbance of running time the time on the test trial minus the time on the first trial of the last day of training. Again, if we consider the mean scores for each group within each study, we find that the range of means for Group H–T was 21.8 to 35.9, while the same figures for Group T–H were from −5.1 to 17.0. A combination of the three one-tail probabilities (.17, .10, and .01) yielded a chi-square value significant at the .01 level.

The results of this experiment are consistent with those of the initial two papers of the present series (5, 6) in the sense that they all lend support to the avoidance hypothesis which has been offered to explain the switching phenomenon. The general implication of the avoidance hypothesis and the switching phenomenon are presented in the two initial papers. It would seem appropriate, however, to make one additional point.

Perhaps the major importance of this series of studies relates to both the empirical and theoretical significance of the construct of the fractional anticipatory goal response. The adherents of the neobehavioristic theory of Hull, Spence, and Miller have been forced to rely upon this construct more and more in their attempts to explain trial-and-error learning. If this construct is to be useful, it must fulfill the criteria of a legitimate theoretical construct. Otherwise it will serve merely as a theoretical escape clause to be invoked when convenient. Brown and Farber (1) have outlined a set of criteria of a legitimate theoretical construct, and we think it would be of interest to demonstrate how the construct of the fractional anticipatory goal response has fulfilled these criteria in this series of studies.

1. A legitimate theoretical construct is not a directly "pointable-at" entity; instead, it is a shorthand expression for a set of relationships. Its only "reality" stems from the empirical predictions which it generates.

The construct of anticipatory goal response is properly viewed as a hypothetical response variable intervening between the directly observable environmental situation and the behavior of *S*. The predictions made in this series of experiments through the use of this construct stem solely from the postulated relationships between this construct and the antecedent conditions and subsequent behavior of *S*.

2. The useful construct is one for which there is a clear statement of its

relations to other concepts in the theoretical framework.

According to our conception, the fractional anticipatory goal response is a hypothetical response which obeys the same principles of behavior as do observable response systems such as bar pressing, salivation, etc. In this sense the construct is integrated with all the other constructs of the neobehavioristic theory on which we have based our analysis. It should be realized, however, that whereas the usual S–R association, such as the classical conditioned salivary response, has a directly observable response element, the association involving an anticipatory goal response has only a hypothetical response element. This requires the behavior theory involving such a construct to make additional assumptions about the influence of this hypothetical response upon overt behavior.

3. Since a theoretical construct is used to predict behavior, the theory must state clearly the relations which hold between the construct and the subsequent behavior of the organism.

In our analysis of the switching phenomenon, we have added to the neobehavioristic system an additional assumption that thirsty *S*s tend to avoid cues conditioned to anticipatory eating responses of dry food. This has permitted our theoretical analysis, which rests heavily upon the construct of the fractional anticipatory goal response, to yield specific deductions about such overt behavior as the selection of a previously nonpreferred side of a T maze and disturbance in running time.

4. To avoid circularity a theory must state the relationship postulated to hold between the construct and the antecedent and the present conditions of the organism.

In our experiments we have proposed relationships between the construct of anticipatory goal response and such variables as the number of trials, nature of the goal object, and type and intensity of deprivation period. Thus we are able to vary independently the strength of the anticipatory goal response by varying any of these.

We therefore conclude that the construct of the fractional anticipatory goal response is capable of being both operationally and theoretically adequate.

Summary

Two groups of *S*s were trained in a T maze in which one of the end boxes was baited. One group was thirsty and received water as a reward while the other group was hungry and received food as a reward. After *S* responded correctly for 12 successive trials its motivational condition was shifted to that of the other group. On the test trial under the new drive state it was found that 41% of the hunger-to-thirst group switched their spatial response, while only 15% of the thirst-to-hunger group switched. This result was consistent with the hypothesis that thirsty *S*s avoid cues conditioned to anticipatory eating responses to dry food.

References

1. Brown, J. S., & Farber, I. E. Emotions conceptualized as intervening variables—with suggestions towards a theory of frustration. *Psychol. Bull.*, 1951, **48**, 464–495.
2. Hull, C. L. Differential habituation to internal stimuli in the albino rat. *J. comp. Psychol.*, 1933, **16**, 255–273.
3. Kendler, H. H. The effect of change of drive upon T-maze performance. *Amer. Psychologist*, 1946, **1**, 246. (Abstract)
4. Kendler, H. H. The influence of simultaneous hunger and thirst drives upon the learning of two opposed spatial responses in the white rat. *J. exp. Psychol.*, 1946, **36**, 212–220.
5. Kendler, H. H., & Levine, S. Studies of the effect of change of drive: I. From hunger to thirst in a T-maze. *J. exp. Psychol.*, 1951, **41**, 429–436.
6. Kendler, H. H., Levine, S., Altchek, E., & Peters, H. Studies of the effect of change of drive: II. From hunger to different intensities of a thirst drive in a T-maze. *J. exp. Psychol.*, 1952, **44**, 1–3.
7. Leeper, R. The role of motivation in learning: a study of the phenomenon of differential motivational control of the utilization of habits. *J. genet. Psychol.*, 1935, **46**, 3–40.

(Received June 16, 1953)

Journal of Experimental Psychology
Vol. 47, No. 3, 1954

RESPONSE TERMINATION OF THE CUE STIMULUS IN CLASSICAL AND INSTRUMENTAL CONDITIONING

DELOS D. WICKENS

The Ohio State University

AND CHARLES E. PLATT

DePauw University

Within recent years several experiments have shown that if a response terminates the eliciting stimulus, this S–R connection is established more rapidly (6) or is made with shorter latency (2, 3) than if the stimulus is terminated some time after the response.

In the study by Mowrer and Lamoreaux, rats were trained to escape from shock in an avoidance-conditioning situation under two conditions. In the experimental condition, escape from the electrified area of the box not only terminated the shock (or actually avoided it if the response occurred before shock onset), but *it also terminated the conditioned stimulus*. In the control condition, shock could be terminated or avoided by the response, but the CS continued for 5 sec. after the response. More rapid conditioning was found for the experimental than for the control animals.

Mowrer interprets this result according to his two-factor theory (5, 6). According to this formulation, the buzzer acquires anxiety-producing tendencies because it is paired with shock. When the response terminates this stimulus it inevitably decreases the tension state that the stimulus evokes. This produces immediate reinforcement of the response. If the tone is not terminated until some time after the response is made, the reinforcement is delayed. If we assume a temporal gradient of the effect of reinforcement, then the behavior in this experiment is easily predicted by Mowrer's theory.

However, the same result could be predicted by Guthrie's nonreinforcement theory (4). According to this theory, a stimulus becomes attached to a particular response if the two occur simultaneously, and this attachment is made in a single trial. This S–R connection will persist unless some conflicting response becomes attached to the stimulus in a similar manner. Ordinarily the correct response serves to act back upon the stimulus situation, modifying or eliminating the stimulus so that an antagonistic response does not become attached to it. This type of immediate change functions for the experimental group in the Mowrer and Lamoreaux experiment, but does so to a much lesser degree for the control.

In an experiment by Brogden, Lipman, and Culler (1), employing a classical design, it was found that, after a certain point in training, the tendency to give the "correct" response decreased with increasing numbers of reinforcements. This was not true of the instrumental situation.

An interpretation in terms of Guthrie's theory accounts for these facts by the assumption that, since the UCS continues in the classical situation, S is likely to give responses which are antagonistic to the correct one. These are made in the presence of the CS and therefore become attached to it. Since they occur later in time than the correct response, they supersede it and these responses, rather than the arbitrary "correct" one, are the responses which are learned. Sheffield's (7) experiment offers some support for this interpretation.

If, however, an experiment were conducted in a classical situation in which the cue or CS was terminated when the correct response occurred, this interfering learning should be avoided. The rate of learning of such a group should differ markedly from another classical group

for which the cue stimulus continues during the occurrence of the shock.

This would not seem to be true for the Mowrer theory. He would probably assume that reinforcement can come from reduction either of the primary (shock) or the secondary (buzzer) source of motivation. Since in the classical situation no reduction in the primary motivation occurs immediately after the response for either group, little difference between experimental and control should be found. Whether or not any difference would occur would seem to depend upon whether or not an immediate partial drive reduction (that of the secondary drive only) increases S–R fixation even though the primary drive is not reduced.

The purpose of the present experiment was to compare the rate of conditioning in a group of *S*s whose response terminated the CS with that of a group whose response did not terminate the CS in both an instrumental avoidance and a classical finger-conditioning situation.

Method

Subjects and apparatus.—The *S*s in this experiment were 122 students of elementary psychology at Ohio State University.

The apparatus consisted of a conditioning board, power sources for producing shock and tone, a circuit for controlling time relations between the CS and UCS, and a circuit and timer for recording response latency.

The conditioning board rested on a table at a convenient height for *S* when seated parallel to it. His arm was restrained on the board by means of flexible straps at the forearm and wrist and a rigid strap across the back of the hand. This rigid strap served the function of an electrode as well as a restraining device. The *S*'s middle finger rested upon a microswitch which, when released, stopped a Springfield timer, thus giving a measure of response latency. In the instrumental-conditioning situation the microswitch also served as the other electrode. In the classical situation a small strip of metal, fastened to the middle finger by a rubber band, served the same function.

The CS was a 1000-cycle tone, produced by a Jackson Oscillator, delivered to *S* by means of earphones. The loudness of the tone was roughly 50 db above threshold.

The level of intensity of the shock was constant for all *S*s in the experiment. The level used had been empirically determined as one which would produce vigorous responses with a latency of about 200 msec. for most *S*s. About 1 *S* in 20 reported the shock too mild to produce effective responses and about 1 *S* in 15 found the shock too strong to be tolerated. These *S*s were not used in the experiment.

Time relations between the CS and UCS were controlled by means of an electronic timing circuit. For the time intervals used, the circuit had an error of $\pm.01$ sec. Response latencies were measured to .01 sec.

The *S* and *E* were in separate rooms. Although *S*'s room was not soundproof and noises from other parts of the building were audible, the various relays were housed in sound-reducing boxes so that noises from the operation of the apparatus could not be heard.

Procedure.—The *S* was brought into the experimental room and was seated in the chair by the conditioning board and the following directions were given:

"This is a conditioning experiment. I do not know how much you know about conditioning, but I will tell you a little about it before you begin. During the experiment, you will hear a tone coming from these earphones which will be placed upon your head, and shortly after that you will receive a shock. The shock comes from this electrode on the back of your hand and from this electrode upon which your finger will be resting. You can break the shock by lifting your finger from the electrode which breaks the circuit. (This sentence was omitted in the classical situation.) During the experiment, you will probably develop a tendency to respond to the tone alone before the shock comes on. This is the conditioned response. You should neither fight against it, nor should you respond voluntarily to the tone. If your finger feels like flying up, let it fly up. But at the same time, do not voluntarily lift your finger to the tone. This may sound like a hard direction to follow now, but I think you will know what I mean when the experiment begins. Remember, you neither respond voluntarily nor fight against responding.

"We are going to begin the experiment by running a few tests to the tone alone and a few tests to the shock alone. After that we will begin by pairing the tone and shock. Are there any questions?"

The *E* then left the room, closed the door, and the following procedure was begun: (*a*) three pretests to tone alone; (*b*) three pretests to

shock alone; (c) twenty-five paired presentations of shock and tone. These pairs were separated by time intervals ranging from 5 to 35 sec. in a prearranged random sequence which was the same for all *Ss* throughout the experiment.

Treatment of groups.—In all groups the tone preceded the shock by 450 msec. For one pair of groups, called *Instantaneous*, the tone was terminated immediately upon the occurrence of the finger response. For another pair of groups, the *Delayed*, the tone was terminated 450 msec. after the shock. Thus, in one pair of groups the response immediately affected the CS and in the other it did not.

For two groups of *Ss*, one Instantaneous and one Delayed, an instrumental conditioning procedure was employed; for two groups a classical procedure was employed, shock continuing for 450 msec. after its onset.

The *N* was 21 for each instrumental group and 37 for each classical group. Six *Ss* were dropped from the experiment for the reasons stated above.

Results

The results of the experiment were considered in two ways. In view of Guthrie's contention that learning occurs in a single trial, the performance of *Ss* following the first training trial was evaluated. In the second method of analysis the total learning period was divided into thirds and the experimental and control groups were compared in each of these sections.

In the instrumental situation, 11 *Ss* of the Instantaneous group gave CR's following the first trial and 2 *Ss* of the Delayed group did so. Using Yates's correction, this produces a chi-square value of 6.887, which is significant

TABLE 1

Mean Number of CR's during Successive Thirds of the Instrumental Conditioning Series

Trials	Instantaneous		Delayed		Differences	
	M	σ	*M*	σ	*t*	*p*
2–9	3.33	3.05	1.67	2.39	1.962	.05
10–17	4.08	3.39	3.54	3.03	.571	.50
18–25	4.54	3.70	4.38	3.56	.156	.80

TABLE 2

Mean Number of CR's during Successive Thirds of the Classical Conditioning Series

Trials	Instantaneous		Delayed		Differences	
	M	σ	*M*	σ	*t*	*p*
2–9	2.95	2.82	2.11	2.53	1.340	.10–.20
10–17	4.34	2.98	3.00	3.00	1.909	.05–.10
18–25	4.66	3.23	3.73	3.03	1.269	.20–.30

($p < .01$). In the classical situation, ten *Ss* responded in both groups. Obviously the chi-square value is zero.

The comparisons for the successive thirds of the learning situation are presented in Tables 1 and 2. The data for instrumental conditioning in Table 1 demonstrate a satisfactory level of significance between the Instantaneous and Delayed groups in the first block of trials, but do not do so after that.

From the data for classical conditioning in Table 2 it is demonstrated that the differences between Instantaneous and Delayed groups, though favoring the Instantaneous, do not reach the usual acceptable level of statistical significance in any block of trials.

Discussion

The results of the experiment would seem to be predicted more readily by the type of theory advanced by Mowrer than by Guthrie's strict contiguity theory.

In the case of the instrumental situation, it would be assumed that for the Instantaneous group, the response not only was reinforced by terminating the shock, but also was reinforced by terminating the tone (thus reducing the anxiety state which had become attached to it). The Delayed group received the drive reduction resulting from shock termination, but not that resulting from tone termination. If the assumption is made that a greater amount of drive

reduction will result in more rapid habit formation, or at least better performance, then the superiority of the Instantaneous group would be predicted.

A similar type of reasoning may be applied to the classical-conditioning situtation. The Instantaneous group received some immediate reinforcement by the termination of the tone; the Delayed group did not. For neither group, however, was the major drive stimulus—that arising from shock—reduced immediately. As a result, differences in the behavior of the two groups would not be expected to be very great.

It is likely, however, that Guthrie's theory could handle the results through its use of the concept of movement-produced stimuli. This concept permits the conclusion that although the external stimulus terminated with the response, the actual cue stimulus did not do so and thus became attached to antagonistic responses.

Regardless of the ambiguity of theory evaluation offered by this experiment, it adds further empirical evidence to support the view that time of termination of the cue stimulus is a determiner of efficiency of performance.

Summary

This experiment was conducted in order to determine the effect of time of termination of the CS on the efficacy of finger conditioning. Both classical and instrumental avoidance-conditioning situations were employed. Within each of these situations the CS was terminated by the response for half the *S*s, and for the other half it continued for 450 msec. after onset of shock. For all groups the tone preceded shock by 450 msec.

Statistically significant support for a superiority in rate of learning was found in the response-terminated group in instrumental conditioning, but was not found in the classical-conditioning group.

The significance of this type of experimental design with reference to Mowrer's and Guthrie's theories was considered.

References

1. Brogden, W. J., Lipman, E. A., & Culler, E. The role of incentive in conditioning and extinction. *Amer. J. Psychol.*, 1938, **51**, 109–117.
2. Chernikoff, R., & Brogden, W. J. The effect of response termination of the stimulus upon reaction time. *J. comp. physiol. Psychol.*, 1949, **42**, 357–364.
3. Chernikoff, R., Gregg, L., & Brogden, W. J. The effect of fixed duration stimulus magnitude upon reaction time to a response terminated stimulus. *J. comp. physiol. Psychol.*, 1950, **43**, 123–128.
4. Guthrie, E. R. Psychological facts and psychological theory. *Psychol. Bull.*, 1946, **43**, 1–20.
5. Mowrer, O. H. *Learning theory and personality dynamics.* New York: Ronald, 1950.
6. Mowrer, O. H., & Lamoreaux, R. R. Avoidance conditioning and signal duration: a study of secondary motivation and reward. *Psychol. Monogr.*, 1942, **54**, No. 5 (Whole No. 247).
7. Sheffield, F. D. Avoidance training and the contiguity principle. *J. comp. physiol. Psychol.*, 1948, **41**, 165–177.

(Received June 29, 1953)

Journal of Experimental Psychology
Vol. 47, No. 3, 1954

THE NATURE OF ERRORS IN EXPERIMENTAL LIE DETECTION

D. VAN BUSKIRK[1] AND F. L. MARCUSE

The State College of Washington

Since the end of World War II, there has been a growing use of lie detection techniques both in private and in government work. Wicker (12) has recently evaluated the history of its use and misuse by the courts as well as its current status, and Cureton (2) has described the attitudes of psychologists and nonpsychologists toward its utilization. Concomitant with the greater use of lie detection techniques comes the need for more research, both "pure" and "applied," i.e., in both the laboratory and the actual field situation (12). The present paper is primarily concerned with the nature of errors that occur in the interpretation of lie detection records.

Procedure

The procedure utilized both the Keeler Polygraph (5) (comprising cardiovascular and respiratory indices) and the Stoelting Galvanoscope (5).[2] It required *S* to lie and made use of 50 male *S*s.

Two stacks (red and black) of playing cards were used as stimuli. There were six cards in each stack. The king and queen in each stack were held by *E* and used as control cards, one to introduce and one to conclude the stimulus series. The remaining four cards from each stack were placed face down on a table. The *S* then selected a card from either of the two stacks (without *E*'s knowledge), memorized it, and returned it to its original position. An examination was then made on the Keeler Polygraph. The stack (red or black) that was used first to question the *S* was determined by chance. The *S* was instructed to say "no" to all questions. He was asked, after each stimulus card, whether he had selected that particular card. The stack was presented twice, i.e., two charts were obtained by use of the Keeler Polygraph. After this the Stoelting Galvanoscope was connected to *S*'s right index finger and adjusted to record zero on the visual indicator. The stack (with the card order which had previously been used with the Keeler Polygraph) was then presented. Four runs were made with the magnitude of response to each card being recorded. After these runs a "bearing up" technique was used, i.e., each stimulus card not responded to on the galvanoscope was dropped until finally only one remained.

The same procedure (using first the Keeler and then the Stoelting) was used with the other stack of cards. In effect the procedure required each *S* to lie in one record and to tell the truth in the other. At the termination of the testing *S* wrote on a form the card he had selected and how he felt during the examination. These forms were not seen until testing and interpretation with all *S*s were completed.

Results

Reliability of Judgment

Before assessing the nature of errors made, the question arises as to how stable such errors were, i.e., would they be repeated? The answer to this question lies in part in the consideration of the validity in correctly selecting cards or records (stacks) as well as in the reliability of interpretation. With regard to validity, the 72% correct selection of cards and the 84% correct selection of records (both significant at the .01 level of confidence) testified to the reliability of interpretation. Nevertheless, a more direct measure of reliability was undertaken. The two *E*s made a judgment about where the lie response occurred, i.e., to which card in each of the 50 records,

[1] This study was done in partial fulfillment of the requirements for the degree of Master of Science in Psychology at The State College of Washington. The author is now at Provost Marshal General's School, Camp Gordon, Ga.

[2] The authors wish to thank Dr. V. A. Leonard of the Department of Police Science Administration of The State College of Washington for making these instruments available.

and one month later made another judgment about the same 50 records. The results indicated 84% agreement on cards and 94% agreement on records between these two judgments. It may be said then that reliability, in this particular lie detection procedure, was satisfactory.

Errors

Fourteen stimulus cards were incorrectly selected[3] and these errors were classified into three types, as discussed below. Then a *post hoc* analysis was made of the objective records to discover, if possible, the source of each error.

Type 1.—In this type of error, the card consistently though incorrectly selected on the basis of two runs was in the wrong stack, i.e., the correct card was in the other stack. There were three errors of this type, as follows: (*a*) Cardiac index was indeterminate. Respiratory index definitely indicated an incorrect card while it gave no specific response to the correct one. Galvanic skin response suggested proper card. This error could be repeated. (*b*) The magnitude of responses for all indices was low. None of the indices consistently indicated the correct card. Whether such an error would be repeated with a more adequate record is not known. (*c*) All indices were specific and clear in pointing to an incorrect card. No responses occurred to the correct card. Phenomenological data revealed nothing. This error could be repeated.

Type 2.—In this type of error, the card consistently though incorrectly selected on the basis of two runs was in the same stack as the correct card but at a different location. This type of error would indicate that *E* can determine that a lie has occurred in a particular record but is not definite as to where the lie is. (It should be noted that in criminal investigation this type of error, i.e., the wrong card in the right record, would not constitute an actual error.) There were four errors of this type, as follows: (*a*) All indices pointed specifically to an incorrect card. No responses were shown to the correct card and phenomenological data were negative. This error could be repeated. (*b*) Cardiac and respiratory indices point to an incorrect card. Galvanic skin response was indeterminate. No specific responses were shown at the correct card. This error could be repeated. (*c*) The cardiac index indicated a specific response to an incorrect card. No responses occurred to the correct card. This error could be repeated. (*d*) Cardiac and respiratory indices indicated a specific response to an incorrect card. Galvanic skin response was indeterminate. No specific responses were shown at the correct card. This error could be repeated.

Type 3.—In this type of error a different card was selected in each of the two runs. The cards thus selected might be at different locations in one stack, or might each be in a different stack. This type of error would indicate that *E* cannot determine consistently either the card or the record in which the lie occurred. There were seven errors of this type, as follows: (*a*) Cardiac index showed a response to an incorrect card. Respiratory index was indeterminate. Galvanic skin response suggested the correct card. It is believed that this error would not have occurred if the galvanic skin response had been given more weight. (*b*) Respiratory index showed a specific response to an incorrect card. The galvanic skin response was indeterminate. The cardiac index was found to show disturbance at the correct card. It is believed that this error might have been avoided by closer analysis. (*c*) All three indices pointed clearly to an incorrect card

[3] The data indicated that *E*s, in their interpretation of records, did not favor a particular stimulus location nor did *S*s, in their selection of a card, favor a particular position in the stack.

and did not show any disturbance at the correct one. This error could be repeated. (*d*) The magnitude of response for cardiac and respiratory indices was low. The galvanic skin response indicated the correct card. Additional records with greater magnitude of response should have been obtained or the galvanic skin response given more weight. It is believed that this error might have been avoided by closer analysis. (*e*) Cardiac and respiratory indices showed specific responses at an incorrect card. The galvanic skin response was indeterminate. This error could be repeated. (*f*) Galvanic and cardiac indices were indeterminate but analysis indicated a disturbance of the respiratory index at the correct card. Cardiovascular response was also low in magnitude. It is believed that this error might have been avoided by closer analysis, or by obtaining a clearer record. (*g*) The cardiac index indicated a specific response to an incorrect card. Other indices were indeterminate. This error could be repeated.

It would appear that 9 or 10 of the 14 errors would be repeated. (While analysis might have been undertaken about the contribution of any particular physiological index to the errors made, it should be remembered that in the present experiment the absolute number of errors was small.)

A second point of interest concerns the types of records in which errors of interpretation are made. To investigate this problem the 50 records were thoroughly mixed and one *E* (FLM), who was unaware of the code number of the error records, judged each record to be either poor or good. Records were judged to be poor if the amplitude of the blood pressure index was very low or if movement resulted in a disjointed record (poor records were still considered to be interpretable). Of the 14 errors made, 7 occurred in the 36 records judged good and 7 in the 14 records judged poor. This would indicate that poor records, as defined above, contribute disproportionately to the errors made and their elimination would reduce the number of errors by 50%.

Discussion

While it is often claimed that the data from "real" life are superior to the "simulated emotion" of the laboratory (11),[4] such claims have not been substantiated (8). Until the superiority of one area can be shown, research in both is required.

In the present study, the percentage of errors is slightly higher, and the percentage of indeterminates (where decisions cannot be reached) slightly lower, than the figures commonly reported (2, 10). In regard to the slightly higher percentage of errors it should be kept in mind that in investigating actual crimes, interpretation of records is made soon after the crime, the investigators are concerned with a specific problem, and they are aware of the existence or nonexistence of other suspects, leads, etc. In short, the investigators are conversant with the total situation. In the laboratory, on the other hand, interpretation of records may often be made a month after they are taken, a number of records from different individuals may be considered at the same time, and the lie tends to be an isolated incident for *S* and not part of a total situation. What is being suggested is not that emotion is present only in performing a crime (8) but that in criminal situations minimal cues are more readily at hand and more likely to be used—often unwittingly (7). When using "cold" records in the laboratory, such minimal cues are less likely to be present and consequently less likely to be used. In regard to the slightly lower percentage of indeterminates it should be kept in mind that *E*, in the experimental situation, is more likely to make a decision when an incorrect decision will not penalize *S*.

If in the present experiment poor records were considered as indeterminates and only good records were interpreted, the validity of card selection would be increased from 72% to 81% and that of record selection from 84% to 92%. At

[4] Chatham, R. B. Personal communication, 1953.

the same time this procedure would have reduced the number of errors by 50%, and the number of *Ss* on whom lie detection techniques could be used would be reduced from 100% to 72%. In a culture where allowing a guilty man to go free is considered a lesser evil than imprisonment of an innocent man, such conservatism is in order.

Related to this question of errors is the problem of the techniques of measuring certain physiological indices. The galvanic skin response has been said by some to be accurate (6) and by others to be inaccurate (2, 3, 4) in both experimental and actual criminal cases. The answer to such opposed views may lie in the technique used. In the present experiment the method of "bearing up" was used with the galvanoscope. In this technique one is able to avoid the uncomfortable arm pressure strain (due to the sphygmomanometer cuff) that accompanies the use of the galvanic response when used in conjunction with the Keeler Polygraph. Consequently, one is able to avoid not only arm pressure but time pressure as well and thus to question in a more thorough and more leisurely fashion. In this procedure, the technique used with the galvanoscope is not made subservient to that required by the measurement of blood pressure on the Keeler Polygraph. While it is true that the method of "bearing up" makes impossible the simultaneous recording of the cardiovascular index, it may be argued that this technique obtains more accurate information from the galvanic skin response.

It has been stated that a specific physiological response to a lie depends upon the presence of an element of fear (4, 9, 11). In an attempt to explore this question, *Ss* were asked to "make comments you may have on how you felt during the interrogation." Twenty-two *Ss* (44%) made some statement. Ten reported that they felt "nervous." (One error was made in this group.) Two commented that they had tried to beat the test by concentrating on another card. (Both records were interpreted correctly.) The remaining ten showed no specific trend and described themselves variously as feeling "curious," "sleepy," "calm," "cold," etc. Phenomenological data, as obtained here, would seem to indicate that some form of apprehensiveness can be engendered in the laboratory.

Summary

Analysis was made of the Keeler Polygraph and Stoelting Galvanoscope records of 50 male *Ss* who were required, by the procedure used, to lie. It was found, reliability and validity being satisfactory, that the 14 errors encountered could be classified into three types. *Post hoc* analysis suggested that approximately 68% of the errors would be repeated. It was also found that poor records, as defined here, contributed disproportionately to the number of errors. Phenomenological data obtained from *Ss* seemed to indicate that apprehensiveness may be present in the laboratory. Discussion centered around the significance of the above findings and the use of different measurement techniques for different physiological indices.

References

1. Cureton, E. E. A consensus as to validity of polygraphic procedures. *Tenn. Law Rev.*, 1953, **22**, 728–742.
2. Inbau, F. E. *Lie detection in criminal interrogation.* (2nd Ed.) Baltimore: Williams & Wilkins, 1948.
3. Keeler, L. Debunking the lie detector. *J. crim. Law and Criminol.*, 1934, **25**, 153–159.
4. Larson, J. A. *Lying and its detection.* Chicago: Univ. of Chicago Press, 1932.
5. Lee, C. D. *An introduction to instrumental deception detection.* Springfield, Ill.: C. C Thomas, 1953.
6. MacNitt, R. D. In defense of the electrodermal response and cardiac amplitude as measures of deception. *J. crim. Law and Criminol.*, 1942, **33**, 266–275.
7. Marcuse, F. L., & Bitterman, M. E. Minimal cues in the peak of tension procedure for determining guilt. *Amer. J. Psychol.*, 1946, **59**, 144–146.
8. Marcuse, F. L. The lie detector and the professor. *Tenn. Law Rev.*, 1953, **22**, 919–921.
9. Marston, W. H. *The lie detector test.* New York: Richard R. Smith, 1938.
10. Trovillo, P. V. Deception test criteria. *J. crim. Law and Criminol.*, 1952, **33**, 338–358.
11. Trovillo, P. V. Scientific proof of credibility. *Tenn. Law Rev.*, 1953, **22**, 743–766.
12. Wicker, W. The polygraph truth test and the law of evidence. *Tenn. Law Rev.*, 1953, **22**, 711–727.

(Received June 1, 1953)

Journal of Experimental Psychology
Vol. 47, No. 3, 1954

THE ROLE OF AUDITORY LOCALIZATION IN ATTENTION AND MEMORY SPAN[1]

D. E. BROADBENT

Applied Psychology Research Unit, Cambridge, England

It is known from the work of Hirsh (5) and Kock (6) that speech is more readily understood in noise when the localization of the speech and noise sources is different. These earlier results, however, do not include the case in which both sources are emitting speech, and the listener does not know in advance which source will be relevant. The importance of this selective function has already been demonstrated (1). Poulton (8) has reported some data on the effect of localization on detection of signals, and the present paper is intended to extend this work. The experiments fulfill two main purposes: Exp. I to examine the relative efficiency of a number of different types of localization, and Exp. II to examine the limits of immediate memory on two separated channels.

Experiment I

Procedure.—The task was that used in the earlier report (1). The *S*s were faced with a visual display and were required to answer such questions as "S-1 from GDO. Is there a heart on position Two? Over." The answer was to be Yes or No, and was preceded by the call-sign of the person asking the question, thus: "GDO from S-1. Yes. Over." Unless both the call-sign and the answer itself were correct, the response was scored as an error. There were two voices asking the questions, each of which kept the same call-sign throughout. From time to time both voices would start questions simultaneously; when this occurred only one of them would begin his message with S-1. The *S* was warned that when this happened he was to ignore the message which did not begin with S-1, and to answer the other. A run on this task consisted of 19 plain messages each on one channel only, interspersed with 12 occasions on which both voices started simultaneously. There were 4 sec. between each message, and the whole run lasted between 3 and 4 min.

The questions were recorded on tape and played back on a two-channel Ferrograph. This was similar to the Model C in performance, but carried two separate tracks on the same tape. Where headphones are mentioned below, S.G. Brown Type D.1 were employed: the loud-speakers were a Grampian G.129 (the "movable" speaker) and that built into the Ferrograph (the "fixed" speaker). Where only one loud-speaker is mentioned, half the *S*s used the movable and half the fixed. The level of playback was 75 db in Groups 1, 2, and 3, and 70 db in Groups 4, 5, and 6 (in which a headset was sometimes used).

Subjects.—The 76 *S*s were all Royal Navy enlisted men under the age of 31. They were divided into six distinct groups, each containing 12 *S*s except for Group 1, which had 16. All groups received a training session on the day before the experiment proper, but this training was only in answering straightforward messages and there were no presentations of two messages simultaneously. Training was given with the type of equipment used in the main experiment: for groups which used two types of equipment half the *S*s were trained with each type. On the day of the main experiment, each group received two different conditions of presentation. One run on the task, as described above, was given under each condition, and there were therefore 12 observations on simultaneous messages from each *S* on each of his two conditions. The order of presentation of conditions was rotated within each group.

Conditions.—We will term *separate* recordings those in which each voice was recorded by itself on one of the two tracks of the recorder. Other *stereophonic* recordings were made from the separate recordings by playing back the latter simultaneously over the two loud-speakers of a similar two-channel recorder and picking up the resulting sounds with two microphones. Thus both voices were recorded on both tracks. There were two subvarieties of the stereophonic recordings, *fused* and *localized*; in the former the recording microphones were each just under 3 ft. from each of the speakers so that the recording of a particular voice was simultaneous on the two tracks. In the latter, one microphone was 1 ft. from one speaker and 3 ft. from the other, while

[1] This work was supported financially by the British Medical Research Council, and thanks are due to the Royal Navy for supplying *S*s and technical equipment.

the second microphone reversed this relation. Thus the recording of one voice was about 2 msec. earlier on Track 1 than on Track 2 while the second voice would be earlier on Track 2 than on Track 1 by the same amount. From the data of Wallach, Newman, and Rosenzweig (12) we would expect the localized recordings, when suitably presented, to produce a perception of two separately localized voices, while the fused recordings would give the same localization for both voices.

Group 1 heard separate recordings played through two loud-speakers. The fixed speaker was kept 3 ft. in front of *S* at shoulder height and facing him, while the movable speaker was at the same distance from *S* and 1.5 ft. higher from the ground. In one run the movable speaker was directly above the fixed one, while in the other it was on *S*'s right-hand side, i.e., separated from the fixed one by 90° in azimuth. The *S*'s head was free.

Group 2 heard stereophonic recordings through two loud-speakers, in the 90° separated position. One run was fused and one localized.

Group 3 heard one track only of the stereophonic recordings, one run being a fused recording and one localized. The track was played through a loud-speaker, half the *S*s hearing Track 1 through the fixed speaker and half Track 2 through the movable speaker in the 90° position.

Group 4 heard stereophonic recordings, one run fused and one localized, each track being led to one ear of a pair of headphones.

Group 5 heard separate recordings in one run through a divided headset with each track to a different ear. In the other run two loud-speakers were used, mounted facing each other level with *S*'s ears on each side of him and with 2 ft. between the speakers.

Group 6 heard separate recordings in one run through the same two-loud-speaker arrange-

TABLE 1

CONDITIONS COMPARED IN LISTENING TO MEANINGFUL MESSAGES
(Experiment I: Only messages for S-1 to be answered)

Group	*N*	Conditions			% Errors		*t**	*p*
		General	1	2	1	2		
1	16	"Separate" recordings coming through two loud-speakers	Speakers one above the other	Speakers 90° apart	30	20	3.30	.01
2	12	"Stereophonic" recording coming through two speakers 90° apart	"Fused" recording	"Localized" recording	58	31	4.49	.01
3	12	As above, only one speaker working						
		(*a*) Voice leading on "localized" recording	"Fused" recording	"Localized" recording	49	36	<1	Insig.
		(*b*) Voice later on "localized" recording	"Fused" recording	"Localized" recording	55	54	<1	Insig.
4	12	"Stereophonic" recordings, each track coming through one ear of a headset	"Fused" recording	"Localized" recording	57	40	3.09	.05
5	12	"Separate" recordings	Each track coming through one ear of a headset	Each track through a speaker opposite one ear	28	16	2.71	.05
6	12	As above	Each track coming through a speaker opposite one ear	One track through an earphone, other through a speaker opposite one ear	19	17	<1	Insig.

* The data required an angular transformation to give normality, and the values of *t* are for the differences between the means of the transformed data.

ment as Group 5. In the other run one speaker was replaced by a single phone mounted alone on the headset and worn on the ear opposite to the remaining speaker.

Analysis of data.—The technique of observing only two conditions on each group of *S*s, rather than that of giving all conditions to all *S*s by latin-square or factorial design, was adopted because interactions of condition and order are suspected to be important in this type of task. Consequently the comparisons intended by the design are all between paired conditions, both members of the pair having been given to the same group of *S*s.

For each comparison the proportion of errors was determined for each *S* in each of his two conditions, corrected for continuity, and transformed to ϕ ($= \sin^{-1}\sqrt{x}$), this being the appropriate transformation for the reasons given by Fisher and Yates (4). Analysis of variance was then carried out, that due to *S*s and to order of presentation extracted, and the variance between conditions compared with the remainder. (Variance due to the use of different equipment in training or, in Groups 3 and 6, to using the particular loud-speaker has of course been confounded by the design with that due to *S*s, and extracted with it.) The final comparison was thus always equivalent to a t test with df equal to two less than the number of *S*s.

The transformed means would be somewhat unintelligible in tabular form, as they are naturally expressed in degrees or radians and have little psychological meaning. For each condition, therefore, the total number of errors is shown expressed as a percentage of the total possible. This should make interpretation easier, but from a logical point of view the t values are the important ones.

Results.—Group 1 shows that the results of Hirsh (5) and Kock (6) apply to the present case as well as to the understanding of speech in noise; i.e., a relevant message presented simultaneously with an irrelevant one is more likely to be understood when the two sources are separated in space.

Group 2 shows that the same advantage applies with apparent differences in localization produced by a stereophonic technique.

Group 3 shows no difference between conditions when listening to only one of the stereophonic channels, which serves as a control for Group 2. It implies that the two types of stereophonic recording did in fact give nearly equal intensities for each voice on both channels.

Groups 2 and 3 together act as a control for Group 1. As in Group 2, the intensity relations between the voices did not alter by an important amount any frequency distortions introduced by room reverberation and should have been the same for both conditions. Therefore the results from Group 1 cannot be explained by changes in such distortions.

Group 4 shows that differences in localization are still useful when produced by stereophonic means through headphones.

Group 5 shows that with this equipment two separate tracks are more easily understood through two loud-speakers than through two separate headphones. This result is, other things being equal, to be expected from Hirsh's (5) data on restricting head movement.

Group 6 showed no difference between a two-loud-speaker system and one using one speaker and a headphone. Together with the results from Group 5 this means that a mixture of two systems of presentation may sometimes be the equal of the better of the two systems.

Experiment II

Throughout Exp. I the percentage of errors was less than 50% in all conditions allowing some spatial separation (i.e., all the two-channel conditions except the fused stereophonic recordings). This contrasts with earlier findings using two voices from one loud-speaker (1) and implies that spatial separation not only favors the understanding of messages once they receive attention, but also that the correct message is more likely to secure attention.

A preliminary experiment was carried out using only call-signs alone, and requiring *S* to pick out a sign from other irrelevant ones without knowing which channel it would occupy. At a slow presentation rate differences in localization were useful and at a fast

rate they were harmful. This suggested the possibility that listeners can hear two spatially separated signals successively rather than simultaneously, and the following experiment on immediate memory was designed to test this hypothesis.

Procedure.—The two-channel Ferrograph was employed, using the split headset with each track delivered to a separate ear. Recordings were played of lists of digits under various conditions described below, and *S* was instructed to listen to each list completely through and then write down as much of it as he could remember.

Subjects.—The 24 *S*s were again Royal Navy enlisted men under the age of 31; 12 of them were assigned to Group 1, and the remainder served both as Group 2 and as Group 3 since the comparisons involved were independent. Six *S*s performed the Group 2 conditions before the Group 3 ones, and six reversed this order. No *S*s had served in Exp. I.

Conditions.—We shall term *conventional* lists those in which a single series of digits was heard identically by both ears, so that *S* heard the series in the normal manner used for determining memory span; e.g., Track 1 = Track 2 = 734215. In *binaural* lists different material was recorded on the two tracks, so that one digit arrived at one of *S*'s ears and another digit simultaneously at the other ear, then another pair, one to each ear, and so on; e.g., Track 1 = 734, Track 2 = 215.

Group 1 heard five conventional lists, each of six digits. A warning was then given, and ten binaural lists given, each consisting of three pairs or six digits in all. A final five conventional lists followed and the whole procedure was then repeated using lists of eight digits instead of six. A ½-sec. interval was used between digits or pairs, and the instructions for the binaural lists were to record the digits in whatever order *S* chose.

Group 2 heard 24 binaural lists, each of three pairs of digits, and *S*s were instructed to record the digits in the actual order of arrival; e.g., Track 1 = 734, Track 2 = 215, Response = 723145. The *E* marked as an error any response in which a digit appeared earlier than one which had arrived before it, but the order of writing each pair was immaterial. The rate of presentation was varied, the first three lists having a 2-sec. interval between pairs, the next three 1½ sec., the next 1 sec., and the next ½ sec. Successive groups of three lists then became slower over the same range.

Group 3 heard the eight-digit lists of Group 1, and in addition heard a series of ten binaural lists, each of eight pairs. In this condition the instructions were to write down the first four digits heard on one ear and the last four heard on the other. For half the lists the first digits to be written were on the right, and for the other half, were on the left. A ½-sec. interval was used throughout. This condition is labeled "control for binaural" in Table 2.

Analysis of data.—As the differences obtained were large, the number of *S*s showing each effect was determined and p calculated from the binomial distribution. For the present case the significances are high even by this method, which does not involve unnecessary assumptions. In Group 2 a t test was also applied to the difference between a 1-sec. rate and a 1½-sec. rate, using the distribution of individual *S* differences. This was because five *S*s showed zero differences and so made the binomial test doubtful, but the t test is of course almost equally suspect.

Results.—Group 1 shows that almost all correct responses to binaural lists were written down in such an order that all digits presented to one ear were written before any on the other ear; e.g., if Track 1 = 723, Track 2 = 215, the response is 723215 or

TABLE 2

CORRECTLY RECALLED LISTS OF DIGITS (EXP. II)

Group	Conditions			% Correct*		
	1	2	3	1	2	3
1 (6-digit lists)	Conventional	Binaural: correctly recorded with all digits on one ear before any on the other	Binaural: correctly recorded in any other order	93	62	3
1 (8-digit lists)	Conventional	As above	As above	56	13	2
3	Conventional	Binaural	Control for binaural	42	5	5

* For Group 1, both lengths of list pooled, $p < .001$ for the difference between Cond. 1–2 and 2–3. All *S*s showed these differences. For Group 3, $p < .001$ for 1–2 and 1–3.

215723. This confirms the hypothesis that when two different simultaneous sounds arrive at the two ears, the sound on one ear produces a response before that on the other ear. In addition, this type of task appears distinctly more difficult than a conventional memory span.

Group 2 shows that failures to record digits in the actual order of arrival are no less frequent at a 1-sec. rate than at a ½-sec. rate, but 11 of the 12 *Ss* show a considerable improvement (Fig. 1) between a 1-sec. and a 2-sec. rate ($p < .01$). The difference between 1- and 1½-sec. rates is due to only seven *Ss* ($p < .01$ by a t test). This seems to imply that when attention is shifted away from one channel to another and then back to the first a time interval of between 1 and 2 sec. will be required.

In addition, the low level of performance at the two fast rates shows that the results of Group 1 are not due to self-instruction by *S*.

Group 3 shows that the binaural and the "control for binaural" presentations give equal increases in difficulty compared with conventional presentation. These two conditions are alike in that in both of them information to be memorized is accompanied by other sounds on the other ear, which may be "distracting"; and also in that a change of channel must be made halfway through the list, which may act as an additional item.

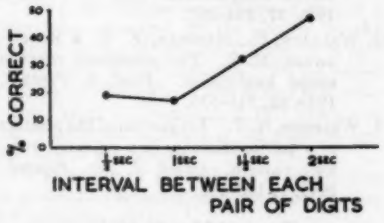

FIG. 1. Recall of binaural spans in the actual order of arrival, as a function of presentation rate (Group 2)

But the two conditions differ in that the binaural one requires the reproduction of information which arrives while attention is directed elsewhere, while the "control for binaural" does not. This difference, which may be regarded as a case of "incidental learning," need not therefore be supposed to produce any extra difficulty.

DISCUSSION

The practical implications of these results may be summarized as follows: some degree of spatial separation is probably desirable in most cases where an individual must deal with two or more channels. But if rapid alternation between two channels is required, separation may cease to be an advantage. This in fact agrees with the results of Poulton (8) on a synthetic representation of the work of control-tower monitors. It is worth noting that the time interval required for double shifts between channels is of the order given by introspective techniques by Pillsbury (7) for double shifts of attention subjectively defined. Other relevant times in the classical work are given by the same author. The interval is also of the same order as the failures which appear in skilled tasks after prolonged performance, failures which have been ascribed on other grounds to shifts of perceptual selection away from the task and back (3). Such time intervals are important for general psychological theory (2).

In addition, the fact that information may be stored temporarily and only later give rise to selective response (attention) is of some interest. It was once well known under the name of "prior entry" (11, 13), and it allows us to harmonize the results on spatially separated speech channels with the view that the perceptual mechanism has a limited capacity (1). But to this view we must add that excess information may be stored. The results are also relevant to the controversy over continuity of learning, particularly the finding that in conditions of equal perceptual difficulty the same number of items can be recalled whether presented successively or simultaneously. It is difficult in this experiment to see any

distinction between memory span and span of apprehension.

These data also place some limits on physiological theories of auditory localization. For instance, Kock (6) stated his findings on the understanding of speech in noise in terms of the insertion by the brain of a time difference between the signals arriving via the two ears. It must be noted that the findings of Exp. II and of the preliminary experiments will support this view only if we suppose that the brain can subsequently remove this time difference and insert a different one, not between future signals, but between past ones. This seems somewhat unlikely, and no doubt Kock intended his phraseology only to emphasize the importance of time as opposed to intensity as a cue, which these results support. If a physiological model is desirable, it should rather be sought in the work of Rosenzweig and Rosenblith (10) and Rosenzweig (9), in which it is shown that time differences at the ears are recorded into spatial position on the cortex. It seems reasonable that information at different cortical points should pass successively through some further mechanism of limited capacity, and that changes in the order of succession or complete neglect of one point should be common.

Summary

In Exp. I *S*s were required to answer messages about a visual display, and occasionally two messages were presented together. When this occurred, only one message, that starting with a particular call-sign, was to be answered and *S* knew that such relevant messages might come from either source. It was confirmed that spatially separated sources led to more correct answers under these conditions, and also that apparent separation produced by stereophonic techniques produced the same result. Various other methods of presentation were also compared.

In Exp. II memory span was determined for lists of digits arriving simultaneously, half on one ear and half on the other. All digits given to one ear were written down by *S* before any on the other; this confirmed a suspicion that spatially separated sounds may pass through the perceptual mechanism successively rather than simultaneously. As the rate of presentation was slowed up, however, it began to be possible to recall lists in the actual order of arrival. Although binaural memory span is more difficult than a conventional one, a control experiment showed that this is due to perceptual confusion rather than to an independent failure of memory.

These results suggest that spatial separation will normally be useful unless alternation between two channels is required at a faster rate than about 1 or 2 sec.; this time interval has been found important in other connections, such as the study of "fatigue."

References

1. Broadbent, D. E. Listening to one of two synchronous messages. *J. exp. Psychol.*, 1952, **44**, 51–55.
2. Broadbent, D. E. Classical conditioning and human watchkeeping. *Psychol. Rev.*, 1953, **60**, 331–339.
3. Broadbent, D. E. Noise, paced performance, and vigilance tasks. *Brit. J. Psychol.*, 1953, **44**, 295–303.
4. Fisher, R. A., & Yates, F. *Statistical tables*. London: Oliver & Boyd, 1948.
5. Hirsh, I. J. The relation between localization and intelligibility. *J. acoust. Soc. Amer.*, 1950, **22**, 196–200.
6. Kock, W. E. Binaural localization and masking. *J. acoust. Soc. Amer.*, 1950, **22**, 801–804.
7. Pillsbury, W. B. "Fluctuations of attention" and the refractory period. *J. Philos. Psychol. sci. Meth.*, 1913, **10**, 181–185.
8. Poulton, E. C. Two-channel listening. *J. exp. Psychol.*, 1953, **46**, 91–96.
9. Rosenzweig, M. R. Representations of the two ears at the auditory cortex. *Amer. J. Physiol.*, 1951, **167**, 147–158.
10. Rosenzweig, M. R., & Rosenblith, W. A. Some electrophysiological correlates of the perception of successive clicks. *J. acoust. Soc. Amer.* 1950, **22**, 878–880.
11. Stone, S. A. Prior entry in the auditory-tactual complication. *Amer. J. Psychol.*, 1926, **37**, 284–291.
12. Wallach, H., Newman, E. B., & Rosenzweig, M. R. The precedence effect in sound localization. *Amer. J. Psychol.*, 1949, **62**, 315–336.
13. Welford, N. T. Estimation of the position of a tactile stimulus in a repeated auditory pattern. *Quart. J. exp. Psychol.*, 1949, **1**, 180–192.

(Received May 18, 1953)

Journal of Experimental Psychology
Vol. 47, No. 3, 1954

AFFECTIVITY AS A FACTOR IN THE APPARENT SIZE OF PICTURED FOOD OBJECTS[1,2]

HOWARD L. BEAMS

University of Pittsburgh

"Perceptual distortion" of need-related objects has been the concern of many recent studies (**1, 2, 3, 4, 7, 8, 9**). In addition to the methodological flaws found in many of these studies which were pointed out by Gilchrist and Nesberg (**4**), they have, almost without exception, reported results based upon differences in the perceived size of various objects between or among groups separated on the basis of inferred needs. It might be assumed that a more fruitful method of investigation would concern itself with intra-individual differences in the perceived size of various objects, when these objects had been demonstrated to have different need properties for the individual. In addition to increasing the precision of the experimental design by allowing the use of each *S* as his own control, this method has the advantage of placing the locus of the investigation in the individual *S*, rather than in the group.

The following study was designed to examine the general hypothesis that affectivity plays a determining role in perceptual organization by testing a specific aspect, namely, the effect of a child's liking or dislike for certain food objects upon his perception of the magnitude of these food objects.

[1] The present paper is a condensed form of a dissertation presented to the faculty of the Graduate School of Syracuse University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author is grateful to Drs. George G. Thompson, R. G. Kuhlen, and Eric Gardner for guidance in the planning and execution of this project.

[2] A summary of this paper was read before the American Psychological Association meeting in Washington, 1952.

METHOD

Stimuli.—Five food objects of the dessert type were selected as stimuli. These were: (*a*) piece of fruitcake (sold in grocery stores in packaged 10-cent slices); (*b*) molasses cookie; (*c*) brownie; (*d*) sugar cookie; (*e*) marshmallow-covered chocolate cupcake, known as a "snowball-cupcake."

After the desserts were selected, a Kodachrome picture of each dessert was taken against a common dull gray background from a distance which would cause the image to be projected life-size at 106 cm. from the viewer. The pictures were taken with a Leica III C F-2 Summitar camera and were then developed into 2 × 2-in. slides, for use in a "Tri-Purpose" projector manufactured by the Society for Visual Education and having a Series 0 5 in. focus Anastigmat projection.

Apparatus.—The same dessert objects which were previously photographed were then mounted individually on pieces of masonite faced with white cardboard. These boards measured 50 by 34 cm. and were designed for use in the presentation of the stimuli in a stimulus box. This was a rectangular box with a wooden frame and masonite walls, painted flat black on the inside to minimize glare, and slate gray on the outside. This box measured 108 × 69 × 46 cm. and contained a light source consisting of a standard silver-colored reflector 22 cm. in diameter and an A-1 Photoflood bulb, as well as a round opening ¼ in. in diameter, through which *S* might view the stimulus monocularly. The end of the box opposite *S*'s position consisted of a door swung on hinges and holding two pieces of grooved wood, by means of which the individual stimulus cards might be attached to the door. The various stimuli could thus be introduced in the stimulus box under standard conditions when needed. When the door of the box was closed and latched, there was no illumination inside the box, thus making it impossible to see the stimulus unless a light were turned on. A "pass-through" switch was rigged to the light source, allowing *E* to expose the stimulus for 5 sec. upon each presentation. The A-1 type Photoflood bulb was selected because the illumination it gave in the stimulus box was most nearly like the illumination presented in the "projection box." Lighting conditions were standard for the presentation of

both the preferred and the disliked objects for each *S*.

The projection box, with over-all dimensions of 156 × 69 × 46 cm., contained a sliding screen of white cardboard measuring 57 × 43½ cm., which was viewed monocularly through a round opening ¼ in. in diameter. This hole was positioned at the same height as the "peephole" in the stimulus box. A small platform at the front of the box held the projector. The screen could be slid a distance of 70 cm., with 68 cm. representing the closest possible point to *S*, and 138 cm. the farthest possible point. A meter stick was attached to the side of the box with a triangular-shaped pointer running to it from the screen platform itself. Thus, an accurate reading in centimeters of the position of the screen was possible after each adjustment.

Subjects.—The *S*s were 60 children (38 girls, 22 boys), ranging in age from 10 yr., 1 mo. to 12 yr., 5 mo., in IQ from 90 to 150 (California Mental Maturity scores), and in visual acuity from 20/20 to 20/30. These *S*s were drawn from an original population of 497 fifth- and sixth-grade children in six public schools in Syracuse, N. Y.

Procedure.—The *S*s were selected on the basis of their likes and dislikes for the desserts used. Mimeographed sheets were prepared which contained a place for *S*'s name and seven straight lines, each one 7 in. long. The midpoint of each line was indicated by a small letter "o." Each *S* was instructed to write the name of his favorite dessert at the extreme right end of the top line and told that this point would represent that dessert. The extreme left end of the line was to represent the dessert which *S* disliked the most. He was then instructed to think of some dessert which he liked only about half as much as his favorite dessert and to let the small "o" at the midpoint of the line represent this choice. This gave each *S* his own "reference line" against which he might check each of the five desserts which were used as stimuli. The *S*s were asked to indicate their liking or dislike for each stimulus object separately by putting an "X" at that point of the line which represented their feeling for the specific dessert. They were given a chance to ask questions, and several examples were given by *E* in order to demonstrate the importance which even ¼ in. or ½ in. of line had in indicating liking or dislike. The list of desserts was read only after *E* was sure that *S*s understood the instructions.

The scale was given on two successive days and then again after two weeks. Since the only function of the rating scale was as a tool to be used for the selection of extremes of liking and dislike, it was decided to accept as judgments only those marks which fell in either the first or last inch of the lines. Also, since the experimental design necessitated the use of *S*s who showed judgments at both extremes, only those children whose marks fitted this criterion were considered as potential *S*s. Since the scale was used to select only extreme judgments, it was not necessary to determine the over-all reliability of the entire scale. It *was* imperative to know whether the *S*s used had made reliable expressions of likes and dislikes for these food objects. This problem was handled in the following manner: It could be assumed that subject A's mark could fall in either of the extreme inches of the line by chance alone on the first presentation. However, it could also be deduced that the probability of subject A's putting his mark in the same extreme inch on the second presentation would be only one in seven, since the line had been separated into 1-in. lengths. If still another presentation were made, the probability that subject A's mark would fall in the same extreme inch that it had on the first two presentations was only 1/49. It was felt that the approximately 2% level of confidence which resulted was adequate for the study.

Sixty *S*s were found who consistently indicated one strong liking and one strong dislike for any two of the food objects. Each *S* was then shown, in the stimulus box, one of the actual food objects. In order to prevent the order of presentation from influencing the adjustments, half the *S*s were shown their disliked dessert first; and the other half, the dessert for which they had indicated strong preference. After *S* had looked at the actual food object from a distance of 106 cm. for a period of 5 sec., he was then asked to adjust the Kodachrome image of the food object (by movement of the sliding screen, enclosed) until it was in his judgment the same size as the actual object. Four such adjustments were made, each following a 5-sec. presentation of the stimulus. A reading in centimeters of the distance of the screen from *S* was taken after each trial. The same procedure was followed with the object for which *S* had indicated opposite affect. The difference between the average of the four "like" readings and the "life-size" reading of the liked object was computed.[2] The same procedure was followed for the disliked object. The

[2] The image was "life-size" when the screen was the same distance from *S* as the actual object had been in the stimulus box. This was arranged by taking the measurements of the stimuli, then matching the images to the measurements. Since all the pictures of the stimuli had been taken from the same distance, all of the objects projected life-size (i.e., the dimensions of the image matched the dimensions of the stimulus) when the screen was positioned 106 cm. from *S*.

TABLE 1

ANALYSIS OF VARIANCE OF "LIKED" AND "DISLIKED" ADJUSTMENT SCORES

Source	*df*	Mean Square	*F*	*p*
Between-likes and dislikes (A)	1	8509.12	61.34*	<.001
			59.75**	<.001
Between-individuals (B)	59	457.68	3.29*	<.001
			3.17**	<.001
Interaction (A × B)	59	176.91	1.27	>.05
Within-cells (C)	360	138.71		
Total	479			
Pooled interaction and within-cells (AB + C)	419	144.08		

* Tested by within-cells.
** Tested by pooled interaction and within-cells.

difference between the two difference scores was then computed and the data subjected to an analysis of variance.

RESULTS

The fact that 51 of the 60 *Ss* had positive difference scores (indicating *S* had made a larger comparative image of the liked dessert than of the disliked food object) gave an early indication of the significant results which were gathered from the complete analysis. The mean difference between "liked" and "disliked" adjustments was found to be 8.54 cm., indicating that the average adjustment for the liked object produced a considerably larger image than did the adjustment for the disliked dessert.

Table 1 shows the mean square between *Ss* and the error mean square (variation within *S*) to differ significantly ($p < .001$) indicating that the tests measured with sufficient accuracy to distinguish between the *Ss*. Use of Jackson's test of sensitivity indicated further that the reliability of measurement was adequate for the study (ρ between .40 and .70).

When errors of measurement[4] were used as a criterion to determine significance of the difference scores, a highly significant *F* resulted ($p < .001$). Also, when the error term used was the combination of error variance of measurement and variability of individual *Ss* (made possible by lack of a significant difference between liked and disliked scores as a function of individual measure) the difference was still highly significant ($p < .001$). The significant ($p < .001$) *F* ratio for between-individuals and the combined error term indicated the liked and disliked measures to be correlated variables.

No significant differences were found when the sample was split into groups on the bases of sex, age, intelligence, and occupational level of parents.

[4] The derived error term was based on both "in" and "out" adjustments and was undoubtedly inflated by this fact.

DISCUSSION

The task assigned *S* was to adjust a picture to that size distance where the pictured object looked most like the real object just seen ("life-size"). "Liked" and "disliked" adjustment scores of the 60 *Ss* differed significantly. It is then safe to infer that the real, liked object must have *looked bigger* than the real, disliked object. Since the same *S*, with all factors such as age, sex, intelligence, and experiental background constant, made both the "liked" and "disliked" adjustments, it seems reasonable to state that the only evident change in *S* during the two types of adjustments was that of affectivity in regard to the stimuli. It would then follow that *S*'s liking or dislike for the food object has influenced his perception of the magnitude of the object. The fact that nine *Ss* had "negative" difference scores (larger pictures for disliked object than for liked object) suggests that there are questions regarding "perceptual distortion" still unanswered.

SUMMARY

The purpose of this study was to examine the hypothesis that affectivity plays a role in perceptual organization. It was proposed that a specific test of this hypothesis might be gained

through attempting to measure the effect which a child's liking or dislike for certain food objects had upon his perception of the magnitude of each food object. The substantiation of the hypothesis is supported by the evidence presented.

References

1. Bruner, J. S., & Goodman, C. C. Value and need as organizing factors in perception. *J. abnorm. soc. Psychol.*, 1947, **42**, 33–44.
2. Bruner, J. S., & Postman, L. Symbolic value as an organizing factor in perception. *J. soc. Psychol.*, 1948, **27**, 203–208.
3. Carter, L. F., & Schooler, K. Value, need and other factors in perception. *Psychol. Rev.*, 1949, **56**, 200–207.
4. Gilchrist, J. C., & Nesberg, L. S. Need and perceptual change in need-related objects. *J. exp. Psychol.*, 1952, **44**, 369–376.
5. Goodenough, F. L. *Developmental psychology.* (2nd Ed.) New York: D. Appleton-Century, 1945.
6. Jackson, R. W. B. Reliability of mental tests. *Brit. J. Psychol.*, 1939, **29**, 267–287.
7. Klein, G. S., Schlesinger, H. J., & Meister, D. E. The effect of personal values on perception: an experimental critique. *Psychol. Rev.*, 1951, **58**, 96–112.
8. Lambert, W. W., Solomon, R. L., & Watson, P. D. Reinforcement and extinction as factors in size estimation. *J. exp. Psychol.*, 1949, **39**, 637–641.
9. Postman, L., Bruner, J. S., & McGinnies, E. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1948, **43**, 142–154.

(Received June 22, 1953)

Journal of Experimental Psychology
Vol. 47, No. 3, 1954

ACCURACY OF RESPONSE AS A FUNCTION OF TARGET WIDTH

EDWARD A. BILODEAU[1]

Perceptual and Motor Skills Research Laboratory, Human Resources Research Center

In almost any perceptual-motor task there is an implicit or explicit definition of the response proficiency required for scoring a hit on a target. Target size is, then, a common variable in motor skills tasks. The research question is whether wide and narrow targets affect behavior differentially, a question posed but not answered in previous studies because of a confusion between the measurement and the elicitation of different behavior. The present study will attempt to answer this question by the method of reporting to *S* that there was associated with the last response an error less than the true error recorded by *E*. The basic assumption in this knowledge of results procedure is that response accuracy is related to the knowledge of errors provided *S*.

Helmick (5) has varied the size of the target in an experiment with the Rotary Pursuit Test. The results, of course, indicated that the larger the target, the greater the time-on-target *score*. Helmick's relation is reasonable because when the target was enlarged, the scoring area was enlarged. Thus, the greater scoring areas included more responses—responses which might have been made by groups with smaller scoring areas but which were left unmeasured. The study by Helmick, then, leaves unanswered the question of whether large and small targets elicit different behavior. The present study, by contrast, deals directly with the problem of the elicitation of different responses by targets of different extents. This is accomplished by measuring the behavior of several target groups with a *common* scoring extent. In other words, the present approach in the rotary pursuit context, for example, would lead to the measurement of performance on a wide and a narrow target with the narrow criterion.

There are two types of targets in psychomotor tasks. In the first type, *S* can see the full extent of the target and its center (the latter may or may not be indicated by some reference marker); in the second type, target extent and center are not discernible at any moment. The psychomotor devices, Rotary Pursuit Test (7), Two-Hand Coordination Test (7), and Pedestal Sight Manipulation Test (7), are examples of the first type, and the manual lever (1) and micrometer (3) are examples of the second type. In the second sort of task, *S*'s signal of response error can be distorted by *E*, because *E*, not *S*, determines what is shown on the display. Once the response is made, the display serves as a representation of the response and of the optimal response. The difference between the two is presumably an index of the amount (and direction) of the error. When *E* controls the display, *S*'s response error can be readily transformed by *E* without the risk of detection from *S*. In the case of

[1] The experimental work for this study was performed as part of the United States Air Force Human Resources Research and Development Program. The opinions or conclusions contained in this report are those of the author. They are not to be construed as necessarily reflecting the views or indorsement of the Department of the Air Force.

tasks such as the rotary pursuit, however, where the target extent and approximate center are immediately discriminable from the nontarget area, the response error is always discriminated by *S* within psychophysical limits. Here the magnitude and direction of the error are difficult to manipulate experimentally because *S* is nearly always aware of the true error.

As a result of previous investigations where target size was manipulated by *E* (devices of the second type), it was tentatively concluded that response accuracy distributions were unaffected by the size of the target used in training (1, 2). In another investigation, moreover, where the true error was immediately discriminable (device of the first type), the conclusion was no different (8). For both types of situations the conclusion was limited by the range of target sizes employed.

Let it be supposed for the moment that a task in which the target extent and center are not seen by *S* is used and that response errors are presented to *S* after being monitored by *E*. If *E* elects to provide *S* with signals of errors of zero magnitude for a wide band of responses, *S*'s target may be said to be wide. Associated with the response band, however, is a center position and a particular response. This single response is characterized by a single magnitude and may be called the target-center, or goal, response. There is a true error of zero associated with the goal response. The preceding development now allows the formulation of the questions to be answered by the present experiment. How is the growth of the goal response associated with target width? To what degree can the error of responding be reduced if *S* is always provided with signals of zero error? This study makes use of targets wider than those previously employed in a task where the target extent and center cannot be seen, and investigates the reduction of response error as a function of target width.

Method

Subjects.—Four hundred naive male *S*s undergoing basic training at Lackland Air Force Base, San Antonio, Texas, were used.

Apparatus.—The apparatus was much like the Manual Lever Test described in previous publications (1, 2). The lever apparatus used in the present study is referred to as Lever B and is fully described elsewhere (4). Three different types of scores are to be used in conjunction with the lever. These are:

Goal score. The numerical score (45) which *S* is asked to achieve on each trial.

True score. *E*'s measure of the magnitude of response.

Reported score. The score displayed to *S* after each response. The reported score is some function of the true score.

Essentially, *S* is required to position a lever in order to obtain the goal score on a display (reported-score) scale in front of him. He is instructed that the farther the lever is pulled the greater is the magnitude of the score to appear on the scale after the completion of the response. The task, then, is to discover the lever position corresponding to the goal score. Obviously, neither the target extent nor the center is visually discriminable before or during practice.

In general, the apparatus consists of three main parts: (*a*) the response lever, (*b*) the display panel, and (*c*) *E*'s control panel.

The lever in its normal resting position is nearly vertically suspended from a shaft passing from the lever side of the apparatus to *E*'s panel. Displacement of the lever causes the shaft to rotate, which in turn displaces a knock-up pointer indicating to *E* the magnitude of response (true score). The length of the lever arm is fixed by placing a hand grip 14 in. from the lever axis. Attached to the lever by a system of pulleys is a weight of 5 lbs. which exerts a force of 10 lbs. against displacement of the lever. The weight acts also to return the lever to its resting position whenever the lever is released.

The *S* is seated before the display panel in such a way that his right hand can be placed upon the grip when the arm is nearly fully extended. A vertical screen prevents *S* from seeing the lever. The reported-score scale and pointer are located behind a window which is 8 or 9 in. away from *S* at eye level. When scores are not being reported, the window is dark and the scale behind cannot be seen. Scores are displayed by *E*'s lighting a lamp behind the window to expose a scale marked off in 100 units

with 0 on the left and 100 on the right. Major divisions are indicated at multiples of 5 and 10. The *S* read the score aloud after its presentation, and all *S*s read their scores correctly.

When the lever is released the knock-up pointer remains at the point of maximum travel, and the true score is read as the point of alignment of the pointer with a scale marked off in 100 units. A vernier allows for reading the response measure in tenths of units. After recording the true score, *E* computes the reported score and then transmits the latter to *S*.

Procedure.—The *S*s were told that the farther back towards the body the lever was pulled, the higher (numerically) the score they would get. They were instructed to pull the lever to a point where they *thought* a score of 45 would be obtained. After the completion of the response, *E* was to display a score for the response just made. If the reported score was higher than 45, the lever had been pulled too far; if the lever had not been pulled far enough, a score of less than 45 would be shown.

The response lever was automatically repositioned at the beginning of each trial. The *S*'s arm remained at his side between responses.

Experimental design.—The *S*s were assigned without bias to the five groups ($N = 80$) and tested individually. The groups were run in a simple counterbalanced order, i.e., 1, 2, 3, 4, 5, 5, 4, 3, 2, 1, etc.

Obtaining a reported score of 45 was defined as the goal for all *S*s. The treatments differed, during the 12 trials of practice, in the constant amount which *E* added to or subtracted from the true score. The constants defining the groups were ± 0, ± 2, ± 4, ± 8, and ± 16. These values (plus .5 for rounding) represent (*a*) the response tolerance allowed for obtaining the goal (45) and (*b*) the amount, provided the goal was not obtained, which was added to the true score if the response was short or the amount which was subtracted if the response was long. For example, for Group ± 16, any true score within the interval 45 ± 16 (plus .5 for rounding) brought about the goal signal (45), but 16 was added to any true score less than 28.5 and 16 was subtracted from any true score greater than 61.5. The transformation rules mean in effect that any reported score was closer to 45 than its true-score counterpart. In summary, then, the five groups differed both in response tolerance allowed for the goal score, and in the amount by which the true response error was reduced whenever the goal score was not reported. In order to allow *E* time to make the transformation, a 5-sec. delay in knowledge of results was standard for all groups.

The duration of the trial cycle was 20 sec., 5 sec. of which were used for a ready period, 3 or 4 sec. for the response, and 5 sec. for the delay in reporting the score. The remaining time was presumably used by *S* in evaluating the consequences of his last response.

Calibration.—With a lever arm of 14 in., the lever was displaced .18 in. in arc or .75° for every unit of the true score. In the previous experiment (1) the group with the narrowest target had to place the lever within a region .94° wide in order to achieve the goal score, whereas, in the widest target group, the lever had to be placed within a region of 4.70°. In the present experiment, however, a much greater range of values was used, the target widths ranging from .75° for the narrowest to more than 24° for the widest. The target widths .75, 3.75, 6.75, 12.75, and 24.75° represent Groups ± 0, ± 2, ± 4, ± 8, and ± 16, respectively.[2]

The true score was recorded on each trial and constituted the sole response measure.[3] It should be obvious that all *S*s were being scored according to a common criterion and were not scored with reference to the definitions of target extent.

Results

For all *S*s the first response was without knowledge of range of lever travel or the relation of reported score to extent of pull. The true-score means of the groups clustered about 45 on Trial 1, dropped to approximately 40 on Trial 2, and thereafter increased regularly to about 45, except for the means of Group ± 16. This group gradually approached 45 throughout the 12 trials, but on Trial 12 was still substantially below the goal.[4]

Response error as a function of target width.—The constant 45.0 was subtracted from each true score in order to obtain the absolute deviation from the goal score (response error). The mean errors are plotted in Fig. 1 as a function of increasing target width in

[2] Data from Lever A showed that differential responses were quickly evoked to two lever positions differing by less than 2° of arc.

[3] The true scores presented in the following section can be converted to degrees by multiplying by .75.

[4] A few *S*s, primarily those of Group ± 16, acted suspicious of the reported score; most of the *S*s, however, appeared to be responding rather impassively during the course of the 12 trials.

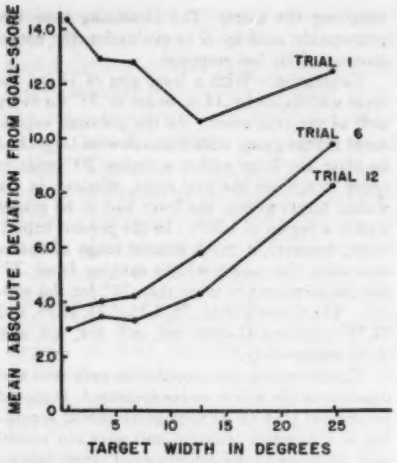

FIG. 1. A plot of mean response error as a function of target width, practice trial as a parameter

degrees; ordinal trial number is the parameter.[5] The upper curve shows that on Trial 1 there was considerable error in responding and a between-groups variability probably associated with sampling fluctuation. On Trial 6, however, the mean error for each group has been reduced, the amount of error reduction for the narrowest target groups being the greatest. Both Trials 6 and 12 suggest systematic between-group differences in that the response error is an increasing, positively accelerated function of target width. Unfortunately, the length of the training session was not sufficient to maintain that the same function is true of asymptotic accuracy.

[5] To save printing costs, tables from which the figures were constructed have been deposited with the American Documentation Institute. Order Document No. 4172 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for 35 mm. microfilm or $1.25 for 6 × 8 in. photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

The trends of Fig. 1 described above were all observed in the *SD*'s which were computed from the measures of absolute deviation.

Thus, the analysis to this point has suggested that (*a*) with successive practice trials the average error of all groups is progressively reduced, (*b*) the rate of error reduction is inversely related to target width, but there appear to be no practical differences between groups over the range .75 to 6.75° of target arc, and (*c*) the preceding summary statements are equally applicable to estimates of variability.

Differential effectiveness of the treatments.—Figure 2 plots the percentage of *S*s who achieved the goal score on successive trials. This figure is presented to show that the values of the experimental variable adequately sampled the range of possible target extents. In this regard, obviously, the treatments were very much differentially effective. For Groups ±0 and

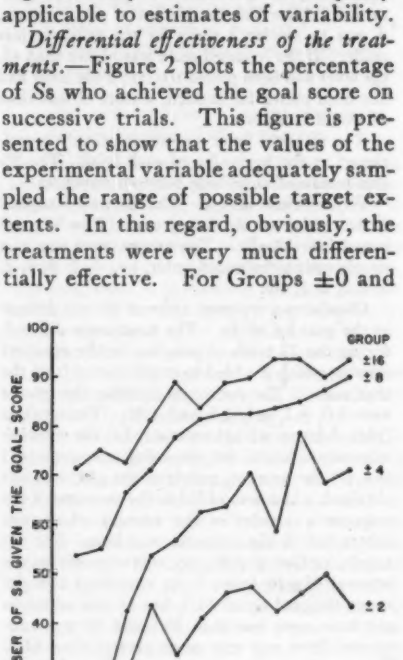

FIG. 2. Percentage of *S*s given a reported score of 45 on each practice trial, target-width group as a parameter

±16, the treatment extremes, about 2% and 71%, respectively, achieved the goal score on the initial trial, whereas the percentages were 15 and 94 on the final trial. Had the range of data for any trial been considerably less (see Fig. 2), the conclusions based on the data represented in Fig. 1 would be restricted to a small range of the experimental variable.

Analysis of within-group phenomena.—A within-group analysis for the treatment extremes based upon *S*s who did or did not receive the goal score would provide information as to the source of error reduction within a group. Such an analysis is suggested by the observation that the mean deviation for Group ±16 by the end of the practice session was still decreasing even though over 90% of the *S*s were receiving reported scores of 45. Too few *S*s, however, of Group ±0 obtained the goal score to make the suggested comparison reliable. Thus, Group ±4 was substituted for Group ±0. The *S*s were assigned to categories on the basis of the magnitude of Trial 1 response errors. The categories of error were, (*a*) errors less than 4.5, (*b*) errors between 4.5 and 16.5, and (*c*) errors greater than 16.5. It follows that for Group ±4, all *S*s in Category *a* received the goal score on Trial 1 and no *S* of Categories *b* and *c* received the goal score. For Group ±16, all *S*s of categories *a* and *b* received the goal score, but all *S*s of Category *c* received information that the response was in error to some degree.

Performance scores for each of the 12 trials for the three categories of Group ±16 are presented in Fig. 3. As a result of the method of classifying the error data, the Trial 1 means are considerably different, but later means tend to converge. The *S*s categorized as receiving some amount of error on Trial 1 reduce their mean deviation from the goal score most quickly and continue the error reduction throughout practice, whereas the *S*s of the two other categories increase the average error of responding from Trial 1 to Trial 2, and only then progressively reduce the magnitude of response error. The figure makes it obvious that from Trial 2 to the end of practice the mean error for all categories was always considerably less than 16.5, or well within the error tolerance allowed for the goal score. An additional analysis revealed that 22 *S*s of Group ±16 remained within their target area on all 12 practice trials. Their performance data gave no evidence of a changing response error, the mean error of Trial 12 being slightly larger than the mean of Trial 1.

Between-category *t* tests were made with the data of Trial 2 to test the hypothesis that the categories of Group ±16 did not differ. The *t*'s were as follows: *a* vs. *b* equals 1.96, *a* vs. *c* equals 2.02, and *b* vs. *c* equals .56 with 55, 36, and 63 *df*, respectively. The test of *a* vs. *b* barely misses the .05 level and *a* vs. *c* is significant at the .05

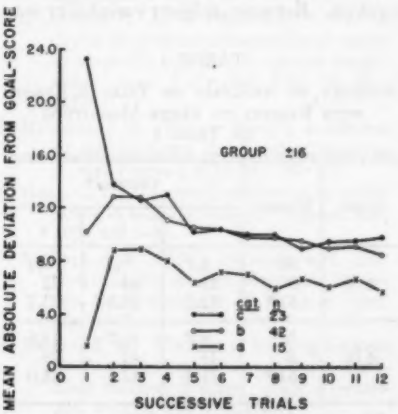

FIG. 3. Mean response error for Group ±16 on successive trials for three categories of magnitude of Trial 1 error

level. It is probable, then, that some between-category differences exist. Other *t* tests for the data of Trial 12 led to the same conclusion.

The analysis of categories for Group ±4 yielded much the same trends as above, except that there were no significant *t*'s and that on Trial 12 the between-category differences were smaller. The data for this group thus suggested that reporting to *S* that he is more accurate (by ±4) than is actually the case produces only short-lived differences between categories of *S*s initially greatly different in error of responding. A summary of the data of Groups ±4 and ±16 is given in Table 1. The entries show that the mean and variance of absolute error in Trial 2 are positively related to the magnitude of error on Trial 1 and that the values of the wider target group are all larger than those of the smaller target group.

Discussion

Evidence was presented to show that informing *S* that his previous response was more accurate than is the case can reduce the speed with which the correct magnitude of response comes to be evoked. Between-subject variability was also positively related to target width. Accuracy of response, over a fairly wide range of distortion of the true scores, appeared unaffected by the transformation, but with fairly severe distortion, accuracy was definitely impaired. The plot of response error as a function of increasing degrees of distortion was increasing and positively accelerated. The lower end of this function showed that beyond a point, *S*s cannot make better use of more precise information. With relatively sizable distortions, less information is transmitted and learning is impaired. It is unlikely that the function is positively accelerated throughout the upper ranges of distortion. For the present task, at least, errors have a maximum, and this maximum suggests an asymptote for the function.

The present experiment verified the conclusions previously proffered with regard to the role of relatively small reported-error distortions (1). The older experiment, as a consequence of its limited range of inquiry, failed to demonstrate the systematic behavioral changes effected by the greater distortions.

Although the analysis of results as presented was not conclusive, there was a tendency for *S*s receiving a goal report on Trial 1 to continue improving with additional trials. For all analyses performed, the mean error of *S*s initially within the target area remained less than one-half the target extent and even tended to decrease with additional practice. Such an analysis does not preclude the possibility of individual *S*s getting outside of the target from time to time, and, as a consequence, obtaining information with regard to the locus of target extremities and so estimating the locus of target center. Under certain circumstances the approximate locus of target center can be guessed to lie midway between responses which are high and those which are low. This process should allow *S* to obtain considerable accuracy in relatively wide target situations. A second analysis of the widest target group, based only upon *S*s who received the goal signal on all trials, failed to demonstrate any learning whatever. Evi-

TABLE 1

Summary of Analyses of Trial 2 Errors with Respect to Error Magnitude on Trial 1

Group	Measure	Category*		
		a	*b*	*c*
± 4	*M*	6.7	9.7	9.2
	N	15	44	21
	SD^2	12.5	28.9	33.2
±16	*M*	8.8	12.7	13.8
	N	15	42	23
	SD^2	31.2	47.5	68.0

* Category *a*, error less than 4.5; Category *b*, error less than 16.5, but greater than 4.5; Category *c*, error greater than 16.5.

dently, knowledge of errors was important for error reduction.

Hull's principles of habit, extinction, generalization, and summation can, perhaps, be used to predict any tendency to center responses as a function of practice (6). When a response is followed by failure to obtain the goal, that response and ones similar tend to be avoided (extinction and a gradient of generalization of extinction). If the goal area is bracketed by responses, gradients of extinction are set up on either side of the goal area. When a response is followed by the goal report, that response and ones similar tend to be repeated (growth of habit strength and generalization of habit strength). With the repetition of several fairly similar responses followed by the goal, the habit strength of any particular response is augmented by a summation effect (summation of habit strength), but the largest habit value after summation is the one most central.

What the present data would indicate is that the growth of the correct response is more quickly established when targets are of fairly small extent (relatively more and greater extinction gradients) than when targets are of fairly large extent (relatively more and greater habit gradients). The data, also, tentatively suggested that if no responses fall outside of the goal area, response accuracy fails to improve. This means that without avoidance processes, there may be no centering of responses based upon the habit summation principle. In order to be more certain of this, however, it is necessary to have assurance that there are no constant errors of response. Were constant errors involved, summation might take place with a habit center away from target center. Thus, even with a summation effect, errors might not decrease.

Summary

The task consisted of learning what magnitude of a lever-positioning response would obtain the goal signal. Five treatments were used, differing from each other in the amount of response error allowed in bringing about the goal signal. This is actually a manipulation of target size. Thus, in all instances it was made to appear that *S* was responding more accurately than *E* was scoring him. The definition of response error, however, was identical from group to group, and was thus independent of the error contained in the knowledge of results.

The data indicated that those groups with the narrower targets learned most rapidly. A plot of magnitude of response error at the end of practice against increasing target size yielded an increasing, positively accelerated function.

Some tentative evidence was presented to suggest that the mean error of *S*s given a perfect reported score on Trial 1 is progressively reduced on subsequent trials, but that there is no error reduction for *S*s who receive a perfect reported score on each practice trial.

The findings were related to *S*s' sources of information, and the relevance of Hullian principles was pointed out.

References

1. Bilodeau, E. A. A preliminary study of the effects of reporting goals as a function of different degrees of response accuracy. *USAF Hum. Resour. Res. Cent., Res. Bull.*, 1952, No. 52–4.
2. Bilodeau, E. A. A further study of the effects of target size and goal attainment upon the development of response accuracy. *USAF Hum. Resour. Res. Cent., Res. Bull.*, 1952, No. 52–7.
3. Bilodeau, E. A. Speed of acquiring a simple motor response as a function of the systematic transformation of knowledge of results. *Amer. J. Psychol.*, 1953, 66, 409–420.
4. Bilodeau, E. A., & Ferguson, T. G. A device for presenting knowledge of results as a variable function of response magnitude. *Amer. J. Psychol.*, 1953, 66, 483–487.
5. Helmick, J. S. Pursuit learning as affected by size of target and speed of rotation. *J. exp. Psychol.*, 1951, 41, 126–138.
6. Hull, C. L. The problem of stimulus equivalence in behavior theory. *Psychol. Rev.*, 1939, 46, 9–30.
7. Melton, A. W. (Ed.) *Apparatus tests.* AAF Aviation Psychology Research Report No. 4. Washington: U. S. Govt Print. Off., 1947.
8. Morin, R. E. Transfer of training between motor tasks varying in precision of movement required to score. *Amer. Psychologist*, 1951, 6, 390. (Abstract)

(Received June 17, 1953)

Journal of Experimental Psychology
Vol. 47, No. 3, 1954

ACCURACY OF RECOGNITION WITH ALTERNATIVES BEFORE AND AFTER THE STIMULUS [1]

DOUGLAS H. LAWRENCE AND GEORGE R. COLES

Stanford University

It has been clearly demonstrated that a subject's accuracy in recognizing tachistoscopically presented stimuli improves when he is told in advance that the object will be one of a limited number of alternatives. There is no common agreement, however, concerning the manner in which this restriction of alternatives facilitates the recognition process. Some investigators speak of "tuning the organism," of "selective sensitization," and of like influences with the seeming implication that the presentation of alternatives modifies *S*'s perception of the stimulus. If the term perception is limited to the actual activity occurring *during the interval of stimulus presentation*, these terms suggest that *S*'s knowledge of what can be expected results in his accentuating those characteristics of the stimulus that are relevant to distinguishing between the alternatives and suppressing those characteristics that are irrelevant to this discrimination. This possibility is strongly suggested by the work of Yokoyama (1) and Chapman (2). Such modification of the perceptual process is to be distinguished from changes that might occur in the memory trace or in the availability of responses as the result of the presentation of alternatives. The present study attempts to evaluate the relative contributions to the accuracy of recognition of these three possibilities: perceptual modification, memory trace modification, and availability of response.

If the major facilitative effect of alternatives is due to the fact that they modify the actual perception during the stimulus presentation, then it follows that this facilitative effect should occur only when the alternatives are given prior to the stimulus, and that it should disappear when the same alternatives are not presented until after the stimulus exposure. Selective perception can operate only when the organism is "set" for the stimulus. If this expected differential effect between alternatives presented before and after a stimulus fails to occur, it rules against the perceptual modification hypothesis.

It would still be possible to argue, however, that two different processes are involved. When the alternatives are presented prior to the stimulus the perception is actually modified, but when presented after the stimulus they facilitate recognition by operating on the memory trace. This two-factor hypothesis can be tested if additional assumptions are made about the relationship between the perception and its memory trace. It seems reasonable to assume that the perception is a "richer," more adequate representation of the stimulus object than is the memory trace. The trace will contain fewer of the stimulus aspects than

[1] This experiment is one in a series of studies of problem solving being done under Project NR 150–104 and supported by Contract Nonr 225(02) between Stanford University and the Office of Naval Research. Work on the contract is under the general direction of Dr. Donald W. Taylor. Permission is granted for reproduction, translation, publication, use, and disposal of this article in whole or in part by or for the United States Government.

the perception, or else contain them in a distorted form.

If this assumption is justified, then discrete alternatives should tend to be equally effective when presented before or after the stimulus, but similar alternatives should have a differential effect under the two conditions. By discrete alternatives is meant those that lead the subject to expect possible perceptual experiences that have little overlap in attributes or characteristics such as color, form, or the like. With such alternatives even a "poor" trace should contain sufficient aspects of the initial perception to permit a discrimination between alternatives. Thus there would be relatively little decrement in the accuracy of recognition when such alternatives are presented after the stimulus rather than before. With similar alternatives, however, there should be a differential effect. By similar alternatives is meant those that lead *S* to expect possible perceptual experiences that overlap in many characteristics and are differentiated on only a relatively few. If such alternatives are presented prior to the stimulus, it is still possible for *S* to accentuate the discriminating characteristics while suppressing the overlapping ones. When presented after the stimulus, however, it is to be expected, on a probability basis, that many of these discriminating characteristics will have dropped out or have been distorted in the trace so that the accuracy of recognition will decrease. Thus in terms of the two-factor hypothesis there should be a measurable interaction between the type of alternative used and whether the alternatives are presented before or after the stimulus.

Method

Subjects and materials.—Sixty undergraduates, divided approximately equally between men and women, were used as *Ss*. The stimuli consisted of 50 black-white pictures of familiar objects cut from magazines. In most instances the background of the picture was homogeneous but in a few it contained shadows, landscapes, and the like. The tachistoscope consisted of an opaque projector with a calibrated camera shutter mounted in front of the lens.

The pictures were divided randomly into five sets of ten pictures each. For each picture two sets of alternatives were constructed, each consisting of the names of four objects including the correct name for the object shown. The first set consisted of discrete alternatives (DA) and the second of similar alternatives (SA). The alternatives were constructed in the following rough manner. The four names in the DA set selected by the *Es* and two judges were names which denoted perceptually discrete objects; i.e., the objects denoted had very little similarity in terms of form, pattern of light and dark areas, and like characteristics when viewed tachistoscopically. In constructing the SA the *Es* were aided by four judges who viewed the pictures tachistoscopically and then listed a series of objects that would appear similar to the picture exposed. The assumption was made that if the stimulus object suggested these alternatives with about equal frequency, they, in turn, when presented to *S* would suggest quite similar perceptual objects. Care was taken in constructing both sets of alternatives to have the correct names for the 50 pictures appear with equal frequency in each of the four possible positions of the set.

Conditions.—The 60 *Ss* were divided randomly into three groups of 20 each. The first group (AB) was given the set of four alternatives just before viewing the picture, the second group (AF) was given the set of four alternatives immediately following the presentation of the picture, and the third group (AN) viewed the pictures but was not given any alternatives.

The reason for including the AN group in the present design was to demonstrate that the sets of alternatives used actually did facilitate the identification of the stimuli. Unless it can be shown that the AB and AF groups are superior to the AN group, it can be argued that *Ss* were not using the alternatives to any extent, that they were relying on recall memory for the identification, and that consequently there would be no reason to expect a differential effect between alternatives presented before and those presented following the stimuli.

Procedure.—The *Ss* were seated at a table 14 ft. from the screen and directly in front of it. A fluorescent overhead light directly in back of them was left on to facilitate the marking of the scoring sheets, to reduce the contrasts in the pic-

TABLE 1

EXPERIMENTAL DESIGN

Subgroups	Exposure Time in Seconds				
	.02	.04	.10	.20	.50
	SA & DA	SA & DA	SA & DA	SA & DA	SA & DA
I	A_1	B_2	C_3	D_2	E_4
II	B_2	C_4	D_1	E_3	A_5
III	C_3	D_5	E_2	A_4	B_1
IV	D_4	E_1	A_3	B_5	C_2
V	E_5	A_2	B_4	C_1	D_3

ture so as to increase the difficulty of identification, and to reduce any possible effect of afterimages. The procedure for Group AB was to project the set of four alternatives on the screen for 30 sec., remove them, and immediately afterwards to expose the picture. As soon as each *S* had marked his paper, the next set of alternatives was presented. In the case of Group AF, the picture was exposed and immediately afterwards the alternatives were presented for 30 sec. Only the pictures were exposed for Group AN, but they were spaced at the same average intervals as for the other two groups. All *S*s were instructed in the appropriate procedure and then shown a demonstration picture tachistoscopically. In addition, the AB and AF groups were shown a demonstration set of alternatives.

Experimental design.—The experimental design is shown in Table 1. Five different exposure times were used, 1/50, 1/25, 1/10, 1/5, and 1/2 sec. The capital letters in the table represent the five different subsets of ten stimuli each that were used. Five of these pictures in each subset were presented with SA and five with DA alternatives. The numbers in the table represent the trials or order of presentation of the stimulus subsets and time intervals. Five *S*s were assigned at random to the five sequences represented by the rows. This design was replicated four times for each group of 20 *S*s. In two of these replications the five stimuli in a subset having SA alternatives in the present table had DA alternatives, and the five having DA had SA. This difference between replications was ignored in the analysis of the data. In the case of the 20 *S*s in Group AN, of course, the SA versus DA was a dummy contrast.

Scoring.—The two groups with alternatives were given scoring sheets on which they checked a number from 1 to 4 depending on which one of the numbered alternatives shown them they thought identified the object. Group AN wrote in the name of the object. These *S*s were instructed always to write in some object even though it involved guessing. The answers of this group were scored correct if, in the opinion of two judges, the name clearly indicated the object even though it was not the specific alternative presented to the other groups. Because of the difference in scoring, no generalizable conclusions can be reached about the relative superiority of recall and recognition memory. Nonetheless, if the AB and the AF groups are superior to the AN group, it supports the contention that in this experiment at least the presence of alternatives actually facilitated the identification of the stimuli.

RESULTS

The accuracy of each of the three groups, as measured in terms of the percentage of pictures correctly identified at each exposure time, is shown in Fig. 1, the uncorrected data in the left panel and the same data corrected for chance in the right. The correction for chance was made by scoring each *S* in the two groups with alternatives on the basis of the number of correct minus one-third the number wrong. This correction is based on the assumption that all the errors made were equally distributed between the four alternatives.

The curves indicate increasing accuracy for all groups with increasing exposure time. Both Group AB and Group AF appear superior to Group AN in this specific situation, but there is no apparent difference between Groups AB and AF. Thus while the sets of four alternatives seem to improve the accuracy of recognition, it apparently makes no difference whether they are given before or after the presentation of the stimulus.

The first statistical check on the data was made to determine if the presence of alternatives before the stimulus significantly increased the accuracy of identification over what would be expected from pure recall

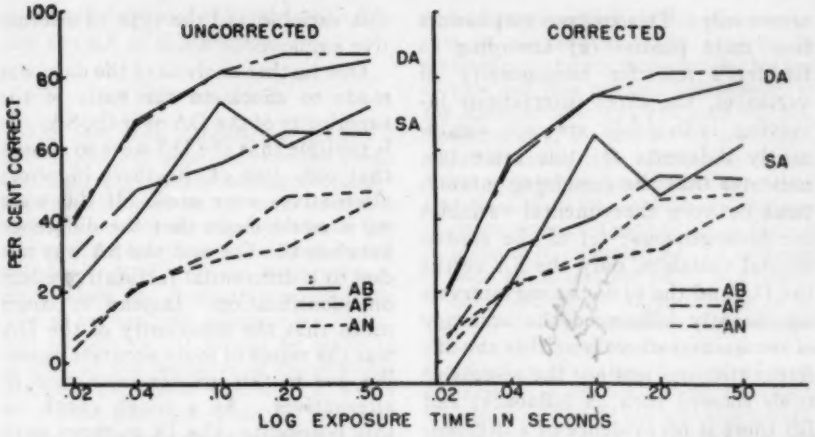

FIG. 1. Each point is based on the judgments of 20 *S*s, five judgments per *S*. The symbols DA and SA stand for discrete alternatives and similar alternatives, and the symbols AB, AF, and AN, respectively, indicate groups with the alternatives presented before and after the stimulus and a group with no alternatives presented. The two AN curves represent a dummy contrast between the DA and SA.

alone. For this purpose each *S* in Group AB was given two scores, the number of SA and the number of DA correctly identified. Each *S* in Group AN was given a score which was one-half the total number correctly identified. The two averages of the AB group were then compared with the average of the AN group by means of *t* tests. In each instance Group AB was superior beyond the .01 level of significance. These tests were then repeated after the scores in Group AB had been corrected for guessing. Again both tests indicated that the latter group was superior to Group AN beyond the .01 level of significance. These results clearly indicate that both the DA and SA used in this experiment improved the accuracy of identification to a greater extent than could be accounted for on the hypothesis of more accurate guessing because of the restriction in alternatives. Thus the evidence suggests that the stimuli and the alternatives used in this experiment are appropriate for a test of the hypothesis that there should be an interaction effect between the type of alternative used and whether the alternatives are presented before or after the stimulus.

The contrast of the DA and SA groups in Fig. 1 indicates that the former permit a more rapid improvement in recognition with increasing exposure time than do the latter, but that this difference tends to be independent of whether the alternatives come before or after the stimulus. It is true that the curves in Fig. 1 suggest that there might be a difference between Groups AB and AF on the similar alternatives, especially at the .1-sec. exposure. But it should be noted that equally large variations are found in the dummy contrast between the two AN groups which experimentally were treated exactly alike.

The statistical significance of these comparisons is confirmed by the analysis of variance presented in Table 2. This analysis compares Groups AB and AF on the basis of uncorrected

scores only. This analysis emphasizes four main points: (*a*) according to Bartlett's test for homogeneity of variances, the three interactions involving individuals are not significantly different; (*b*) this same test indicates that the remaining interactions between experimental variables are homogeneous; (*c*) of the experimental variables, only the SA versus the DA and the various time intervals significantly influenced the accuracy of recognition whereas neither the different stimulus sets nor the successive trials showed such an influence; and (*d*) there is no evidence of a differential influence between alternatives given before and those given after the stimulus or of an interaction between this variable and the type of alternative employed.

TABLE 2

ANALYSIS OF VARIANCE OF UNCORRECTED SCORES FOR GROUPS AB AND AF

Source	*df*	Variance
1. Alt. before vs. after	1	0.25
2. Discrete vs. similar alt.	1	86.49*
3. Time	4	69.29*
4. Stimulus sets	4	2.58
5. Trials	4	3.93
6. Subgroups	4	2.12
7. Confounded interactions		
a. (3 × 4 × 5 × 6)	8	1.73
b. 1 × (3 × 4 × 5 × 6)	8	0.39
c. 2 × (3 × 4 × 5 × 6)	8	1.07
d. 1 × 2 × (3 × 4 × 5 × 6)	8	0.48
8. Unconfounded interactions		
a. 1 × 2	1	0.36
b. 1 × 3	4	1.78
c. 1 × 4	4	0.77
d. 1 × 5	4	1.03
e. 1 × 6	4	0.81
f. 2 × 3	4	0.55
g. 2 × 4	4	1.37
h. 2 × 5	4	3.50
i. 2 × 6	4	4.08
j. 1 × 2 × 3	4	1.59
k. 1 × 2 × 4	4	0.31
l. 1 × 2 × 5	4	0.33
m. 1 × 2 × 6	4	1.73
9. Individuals treated alike	30	1.67
a. 2 × 9	30	0.94
b. 9 × conditions	120	0.87
c. 2 × 9 × conditions	120	0.82

* Beyond the .01 level of significance.

One further analysis of the data was made to check on the basis of the superiority of the DA over the SA. It is possible that the DA were so chosen that only two of the three incorrect alternatives were used. If this were so, it could mean that the difference between the DA and the SA was not due to a differential facilitative effect on identification. Instead it could mean that the superiority of the DA was the result of more accurate guessing due to this greater restriction of alternatives. As a rough check on this possibility, the 18 pictures were selected on which the total errors for the combined AB and AF groups were five or more on the DA and five or more on the SA. Using the distribution of errors between the three wrong alternatives, two measures of uncertainty (3) were computed for each picture, one for the DA set and one for the SA set. This measure has the value of 1.58 when the errors are equally distributed between the three wrong alternatives, and decreases as the distribution becomes unequal. The weighted average of these 18 uncertainty measures for the DA was 1.18 and for the SA it was 1.32. In each instance the average value indicates some discrepancy from perfect equality of distribution, but it is doubtful if these discrepancies are much larger than would be expected by chance for such small samples of errors. Similarly, while the mean of the DA is somewhat smaller than that of the SA, only 12 of the 18 comparisons showed a difference in this direction while the remainder showed it in the opposite direction. Thus while the evidence is very tentative, it suggests that the errors tended to be distributed evenly between the three incorrect alternatives and that there

was no real difference between the DA and the SA in this respect.

Discussion

The fact that alternatives presented both before and after the stimulus facilitate the recognition of tachistoscopically presented stimuli to an equal extent is clearly contrary to any hypothesis implying that perception is modified during the interval of stimulus presentation. Thus when the alternatives are said to "tune" or "set" the organism for "selective perception," it cannot be taken to mean that the reception of the stimulus complex is different from what it would be if those alternatives had not been presented. Rather it must be interpreted as meaning that the facilitative effect of alternatives probably results from their interaction with the memory trace. It is conceivable that even when the alternatives are presented prior to the stimulus, their actual influence is checked until after the reception of the stimulus event, especially with tachistoscopic exposures. An alternative to this emphasis on the memory trace is that the influence of alternatives is primarily on the response variables in the sense of making available or facilitating the occurrence of responses that otherwise would not be made.

These latter two hypotheses are made more plausible by the finding that there is no measurable interaction between the type of alternative used and whether the alternatives are presented before or after the stimulus. If taken at its face value, this clearly rules out the possibility that alternatives presented before the stimulus modify the perceptual process but those presented afterwards operate upon the memory trace. It should be emphasized that the stimuli and alternatives employed in this experiment appear to be quite appropriate for a test of this interaction hypothesis. The initial statistical analysis indicated that in Group AB both the DA and the SA facilitated identification to a greater extent than could be accounted for on the assumption that only recall memory was involved and that the greater accuracy was due to improved guessing as the result of a restriction in alternatives. The analysis in terms of uncertainty measures also tended to rule out this possible explanation of the superiority of the DA over the SA. Consequently, if there were a real interaction between type of alternative and whether they were present before or after the stimulus in this experiment, the sets of alternatives used were sensitive enough to measure it. Despite this, no evidence of such an interaction effect was found.

The generality of these findings is restricted by the present experimental design. It is possible that an increase in the number of alternatives beyond four, an increase in the time interval between the alternatives and the stimulus presentation, or even a more sensitive index of what is perceived would show a differential in favor of the alternatives presented prior to the stimulus. It should be noted, however, that even the present emphasis on the role of the memory trace would predict a differential in favor of alternatives presented prior to the stimulus when the temporal interval between the stimulus and alternatives is lengthened. With the prior alternatives, there is little reason to anticipate that they will be forgotten over relatively long intervals. Consequently, their full influence on the memory image can take effect immediately after the presentation of the stimulus. With alternatives presented after the stimulus, their influence on the trace cannot take place until after the delay interval. During this interval, however, there should be a progressive forgetting of the trace. In such circumstances it is to be expected that alternatives presented after the stimulus would be less effective than those presented prior to the stimulus. It is only when the interval is very brief, as in the present study, that strict equality would obtain. A similar argument applies to an increase in the number of alternatives. Increasing the alternatives implies that the number of discriminating characteristics between them become more restricted, and thus are more likely to be lost in the trace. With these further implications in mind, an even stronger argument can be made for

the parsimony of the present hypothesis that the alternatives facilitate the recognition process by operating on the memory trace or the responses rather than by modifying the perception itself.

Summary

Sixty *Ss*, divided into three groups of 20 each, were shown tachistoscopically presented pictures at various exposure intervals and asked to identify them. The control group had no alternatives presented to it, whereas the two experimental groups each had a set of four alternatives specified for each stimulus. One experimental group had the alternatives presented before the stimulus exposure and the other group following the exposure. The alternatives were of two types: (*a*) SA, or very similar alternatives all suggesting perceptually similar objects, and (*b*) DA, or discrete alternatives all suggesting perceptually different objects. The results and the interpretations made were as follows:

1. The two groups with alternatives were superior to the control group without alternatives in accuracy of recognition even when allowance was made for guessing.

2. Alternatives before and after the stimulus facilitated recognition to an equal extent.

3. The facilitative effect of DA was consistently greater than that of the SA at all exposure times, but there was no evidence of an interaction effect between the type of alternative and when it was presented in relation to the stimulus exposure.

4. These results tend to rule out the hypothesis that the facilitative effect of alternatives results from an actual modification of perception and support the hypothesis that they operate either on the memory trace or response aspects of the recognition process.

References

1. Boring, E. G. Attribute and sensation. *Amer. J. Psychol.*, 1924, **35**, 301–304.
2. Chapman, D. W. Relative effects of determinate and indeterminate *Aufgaben*. *Amer. J. Psychol.*, 1932, **44**, 163–174.
3. Newman, E. B. Computational methods useful in analyzing series of binary data. *Amer. J. Psychol.*, 1951, **64**, 252–262.

(Received for early publication November 4, 1953)

Journal of Experimental Psychology
Vol. 47, No. 3, 1954

THE ROLE OF PROPRIOCEPTIVE FEEDBACK IN POSITIONING RESPONSES [1]

BERNARD WEISS

University of Rochester

Motor responses must sometimes be executed under conditions that deprive the individual of immediate visual confirmation of the effects of the response. Such conditions arise when the visual feedback is wholly absent, or distorted, or lags behind the motor response by an interval of time so long that it prevents the subject from using this feedback to guide his response. When this occurs, the unreliability of the external (visual) feedback loop forces the subject to monitor his responses via his internal (proprioceptive) feedback system. The experiment reported here was an investigation of motor performance under these conditions.

The specific aim of the present study was to determine how the accuracy of positioning responses varies as displacement and pressure information are varied. This information consisted of the changes of muscle tissue in tension and in length that resulted when a subject moved a control. Systematic variation of these changes was achieved by varying the force and extent of movement needed to perform a task in which the visual effect of the movement was delayed.

Previous studies have dealt with various aspects of this problem. Gibbs and Baker (5) and Gibbs (4) compared performance on an isometric pressure control to performance on a free-moving control. Single correction and continuous tracking tasks were used which involved rate tracking, so that movement of the control produced velocity changes. Pressure control was found to be superior in reaction time and movement time in the single correction task and to yield less error in a continuous tracking task.

Jenkins (7) and Hick (6) have studied the discrimination of pressure with isometric controls. Jenkins found that the constant error tended to be positive and relatively great at low pressures, and negative and relatively less at high ones. Hick studied the accuracy with which incremental and decremental forces could be applied to a basal force and found overshooting in both directions for the relatively low pressures used.

The discrimination of distance has been investigated by Brown, Knauft, and Rosenbaum (1) and Spragg, Devoe, and Davidson (9). Both groups found relative constant error and variability to decrease with distance.

A practical aspect of the present experiment is the problem of control "feel," or the optimal force-displacement relationship of a control. A review of this problem may be found in a paper by Orlansky (8).

METHOD

Two sets of force-displacement conditions were used. In one set, the maximum pressure required on the stick was kept constant at 30 lb. and the maximum angular displacement was varied from 3° to 30°. A set of six displacement

[1] Based on a thesis submitted in partial fulfillment of the requirements for the Ph.D. degree to the Department of Psychology, University of Rochester. The writer would like to express his appreciation to Professor S. D. S. Spragg for his guidance throughout the investigation. This study was conducted as part of a research program on human motor performance supported by a contract between the Special Devices Center, Office of Naval Research, and the University of Rochester.

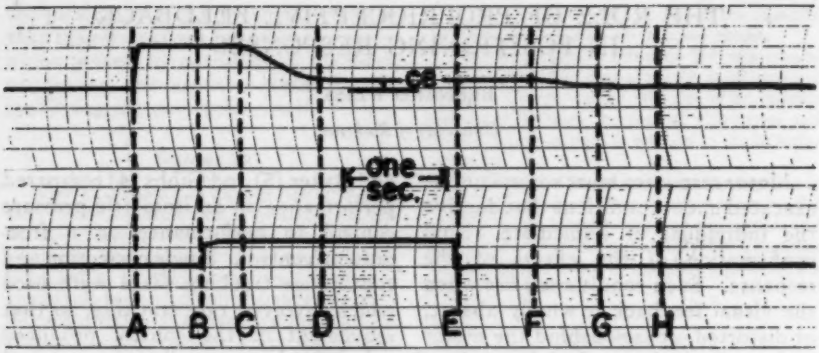

FIG. 1. Sample record

ranges was obtained in this manner. These were: 0–3, 0–6, 0–12, 0–18, 0–24, and 0–30 degrees.

In the other set, the maximum angular displacement was kept constant at 30°, and the maximum pressure required on the stick was varied from 0 to 30 lb. in 6-lb. steps. This yielded pressure ranges of 0, 0–6, 0–12, 0–18, 0–24, and 0–30 pounds. A total of 11 conditions, therefore, were used, with the 0–30 lb., 0°–30° condition common to both sets.

In all conditions, the angular displacement and pressure were linearly related. For each pressure-displacement condition, four displacements corresponding to 25, 50, 75, and 100% of the maximum displacement were used as the required setting in both the fore and aft directions. These corresponded to .35, .80, 1.25, and 1.70 in. displacement on the visual display. The effective length of the stick was 20 in.

Apparatus.[2]—A detailed description of the apparatus appears in (10). The *S* sat in an aircraft-type seat facing the screen of an oscilloscope. His task was to compensate for the vertical displacement of a spot of light from the center of the screen by moving an airplane-type control stick the proper distance and in the proper direction. Fore and aft movements only were possible.

Movement of the stick produced two results. First, it rotated two potentiometers, one in the oscilloscope and one in the recording circuit. Second, springs enclosed in a spring assembly attached to the stick were compressed. By choosing springs of appropriate rates, the pressure variation requirements for the different force-displacement conditions were fulfilled.

The task.—The procedure adopted compelled *S* to respond on the basis of learned proprioceptive cues, and can be illustrated with reference to the sample record shown in Fig. 1, which is an analogue of what appeared on the screen. The spot of light appeared at *A* at the level of one of the eight visual displacements (four in each direction) and disappeared at *B*. *AB* is equal to .6 sec. The interval *BE*, the "blackout" period, is 2.5 sec. and during this time *S* moved the stick the distance and in the direction that he expected would return the spot to the center of the screen. His task was thus a predictive one. He was instructed to move the stick smoothly to the desired setting and to hold it there until the spot reappeared. In the sample record *S* began his response at *C* and ended it at *D*, so that *CD* represents his movement time. The response form is typical of that obtained with practiced *S*s. The spot reappeared at *E* for a 2-sec. period, which permitted *S* to correct for error. In Fig. 1, the corrective response began at *F* and ended at *G*. *H* signifies the end of the correction period. At this time the spot was withdrawn for 5 sec. During this interval *S* moved the stick back to the vertical position and waited for the next displacement to appear. The sequence and duration of the presented events were presented automatically by a selective sequence programmer and timer and were recorded, along with *S*'s response, on a Brush BL-202 inkwriter oscillograph.

Subjects.—All 11 *S*s were male graduate students or staff members in the psychology department of the University of Rochester and ranged in age from 24 to 31 yr.

Design.—Each *S* contributed data for each of the 11 force-displacement conditions. The

[2] The writer would like to thank Mr. Gordon Gay, instrument maker in the University of Rochester psychology department, for his many contributions to the design and construction of the equipment.

effects of sequence of conditions were controlled by a latin-square design.

Procedure.—Prior to the formal sessions, *S* appeared for two orientation and practice sessions. The first was used to introduce him to the equipment and the task, and the second to provide him with further practice. The purpose of these two sessions was to bring *S* to a level of competence on the task high enough to avoid significant variance due to learning after the formal sessions began. Data collected in these sessions showed that the steepest decrease in error for performance on a particular condition occurred during the first trial of 32 settings (4 each on each of the 4 push and 4 pull settings). Later trials displayed no significant decrease in error, and lay on the asymptote of the learning curve.

In the formal experimental sessions, the 8 displacements, 4 in each direction, were first flashed on the screen manually, so that *S* could become acquainted with the force and extent requirements of the condition to be used in that session. The timer was then switched on and the first trial began. Each trial consisted of 32 displacements and corresponding settings to be made. Four settings had to be made for each of the 4 push and 4 pull displacements. Each experimental session consisted of 4 trials, so that in one session *S* made a total of 128 settings, or 16 for each displacement.

Results

The results are based on the last trial of each session. Each of the 8 displacements within a condition was defined by data from 44 settings, or 4 by each *S*. For each setting two scores were obtained from the record, as shown in Fig. 2: error and movement time. The error is denoted by *ce* and is the distance between the line that represents *S*'s response and the zero line of the record (perfect compensation). It was measured at *E*, the end of the "blackout" period. Movement time is the distance *CD* and was measured, but the data are not reported here because they are not meaningful in this context.

CE as a function of displacement.—This section gives the mean constant errors (CE) and their *SD*'s. All these values are expressed as a percentage of the displacement, i.e., as relative error, and all computations were performed with relative error values as the scores. Mean constant error expressed as a percentage of the displacement is given in Table 1 for the conditions where pressure range remained constant at 0–30 lb. while distance range varied from 0–3° to 0–30°. These values are based on the 44 scores obtained for

TABLE 1

Mean Percentage CE as a Function of Extent Range, Pressure Range Constant at 0–30 Lb.

Displacement in Pounds	Extent Range					
	0–3°	0–6°	0–12°	0–18°	0–24°	0–30°
Pull						
7.5	51.91	35.67	18.45	24.52	23.41	13.42
15	5.67	9.30	6.73	2.31	5.57	3.23
22.5	−3.63	−2.24	−4.24	−5.08	−9.18	−3.18
30	−0.95	2.67	−0.16	−4.98	−5.66	−3.94
Mean for pull	13.25	11.35	5.20	4.19	3.53	2.38
Push						
7.5	26.15	8.70	21.80	11.46	7.90	5.82
15	2.92	−4.83	6.90	5.54	1.22	−3.19
22.5	−3.01	−9.42	2.06	−4.66	−0.49	−2.50
30	−5.93	−3.34	2.35	−2.26	−1.39	−5.15
Mean for push	5.03	−2.22	8.28	2.52	1.81	−1.26
Mean for push and pull	9.14	4.56	6.74	3.36	2.67	0.56

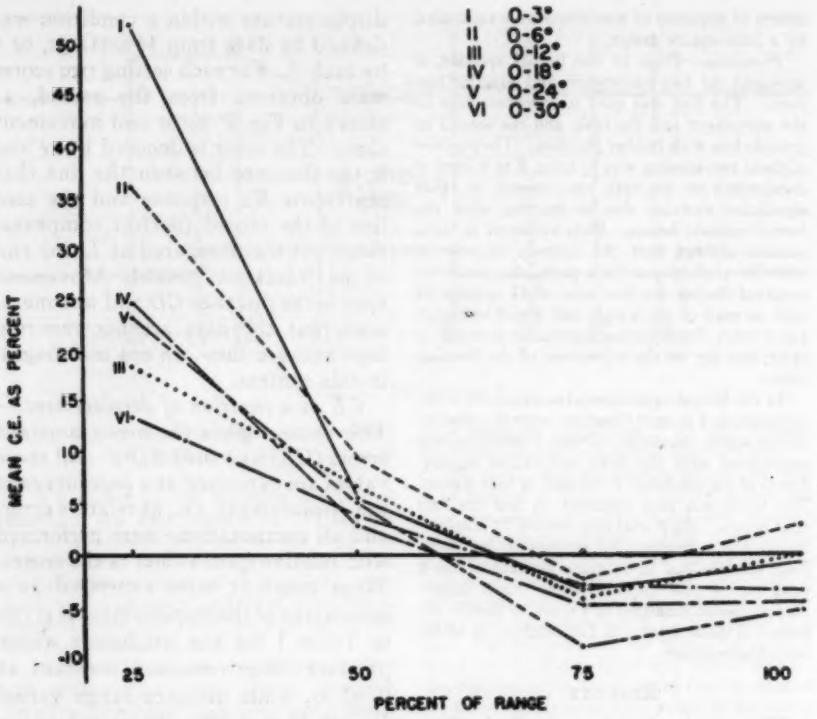

FIG. 2. Mean percentage CE's as a function of percentage of displacement range, maximum pressure constant at 30 lb. (pull)

each displacement within a condition. When the scores from the four displacements within a condition are combined to give mean percentage CE for the condition as a whole, the CE's for pull show a consistent decline, steep at first and then shallower, as the range of extent is increased. The CE's for push are not as regular. A regression equation indicates that this trend, for push and pull combined, is significant beyond the .01 level of confidence.

Figures 2 and 3 are plots of these six conditions in terms of the mean percentage CE for the four displacements in each direction. Figure 2 is for pull and Fig. 3 is for push. On the abscissa the four displacements for each condition are expressed as a percentage of the range of displacement, i.e., as a percentage of the maximum displacement of each condition.

Both the pull and the push data display the same trend. The greatest relative error occurred at the smallest displacement. The other three mean CE's are essentially equal to zero. For both directions the absolute difference in mean CE between the first and second displacements, for all six conditions combined, is significant, by the t test, beyond the .01 level of confidence. Also, the mean CE at the 25% displacement of the 0–30° displacement range is significantly smaller than the mean CE at the 25% displacement

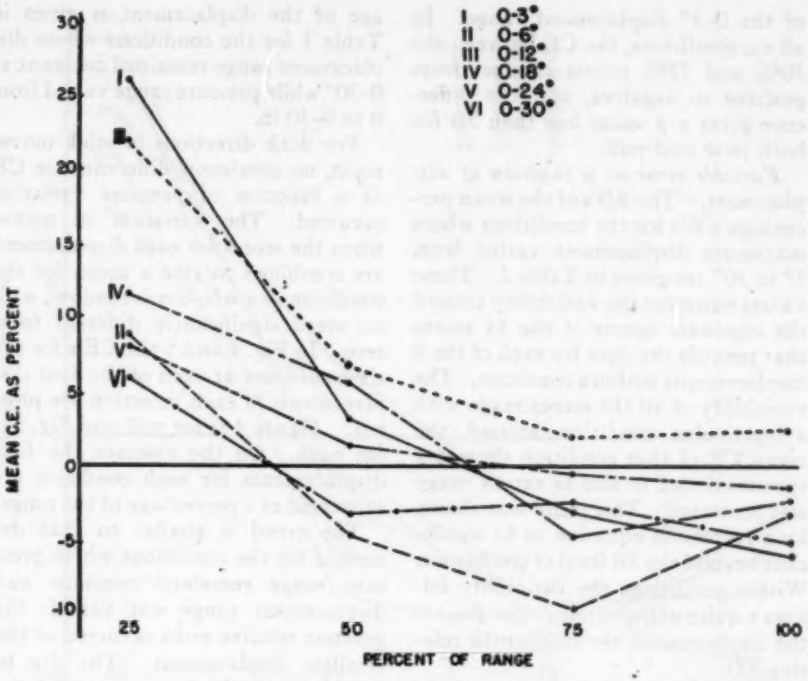

FIG. 3. Mean percentage CE's as a function of percentage of displacement range, maximum pressure constant at 30 lb. (push)

TABLE 2

SD OF PERCENTAGE CE AS A FUNCTION OF EXTENT RANGE, PRESSURE RANGE CONSTANT AT 0–30 LB.

Displacement in Pounds	Extent Range					
	0–3°	0–6°	0–12°	0–18°	0–24°	0–30°
Pull						
7.5	54.66	47.78	27.68	37.56	30.26	31.01
15	27.11	28.94	22.20	18.82	16.26	23.66
22.5	17.12	19.63	23.16	15.96	12.89	12.66
30	15.05	15.14	15.30	12.49	12.26	9.34
Pull (combined)	39.63	33.88	24.11	26.28	23.14	22.15
Push						
7.5	31.51	31.62	25.82	27.72	18.85	20.65
15	29.72	17.18	24.04	15.13	19.92	15.70
22.5	19.01	16.44	18.16	13.26	12.72	12.80
30	20.84	17.11	15.60	10.01	9.53	11.89
Push (combined)	28.74	22.57	22.81	18.94	16.30	16.20
Push and pull (combined)	34.86	29.57	23.52	22.93	20.03	19.49

of the 0–3° displacement range. In all six conditions, the CE between the 50% and 75% points changes from positive to negative, and the difference gives a p value less than .01 for both push and pull.

Variable error as a function of displacement.—The *SD*'s of the mean percentage CE's for the conditions where maximum displacement varied from 3° to 30° are given in Table 2. These values represent the variability around the algebraic means of the 44 scores that provide the data for each of the 8 displacements within a condition. The variability of all the scores made with a particular condition around the mean CE of that condition showed a systematic fall in size as extent range was increased. This trend was shown by a regression equation to be significant beyond the .01 level of confidence. Within conditions the variability follows a quite uniform trend: the greater the displacement, the smaller the relative *SD*.

CE as a function of pressure.—Mean constant error expressed as a percentage of the displacement is given in Table 3 for the conditions where displacement range remained constant at 0–30° while pressure range varied from 0 to 0–30 lb.

For both directions of stick movement, no consistent differences in CE as a function of pressure variation occurred. The variation in means when the scores for each displacement are combined to give a mean for the condition as a whole was random, with no mean significantly different from zero. In Fig. 4 and 5 the CE's for the six conditions at each of the four displacements in each direction are plotted. Figure 4 is for pull and Fig. 5 is for push. On the abscissa the four displacements for each condition are expressed as a percentage of the range.

The trend is similar to that described for the conditions where pressure range remained constant and displacement range was varied: the greatest relative error occurred at the smallest displacement. The dip in error between the first and second displacements for the six conditions com-

TABLE 3

MEAN PERCENTAGE CE AS A FUNCTION OF PRESSURE RANGE, EXTENT RANGE CONSTANT AT 0–30°

Displacement in Degrees	Pressure Range					
	0 Lb.	0–6 Lb.	0–12 Lb.	0–18 Lb.	0–24 Lb.	0–30 Lb.
Pull						
7.5	15.59	17.18	9.31	16.47	11.49	13.42
15	−1.44	−2.62	−1.39	−6.32	−4.62	3.23
22.5	−7.62	−6.82	−6.13	−8.82	−8.10	−3.18
30	−4.82	−6.33	−0.05	−3.47	−9.01	−3.94
Mean for pull	0.42	0.35	0.43	−0.54	−2.61	2.38
Push						
7.5	13.02	6.27	8.56	9.51	3.51	5.82
15	7.66	2.24	1.02	2.77	3.91	−3.19
22.5	0.28	−4.55	−3.15	2.75	−3.60	−2.50
30	−2.39	−4.11	−6.30	−1.35	−4.08	−5.15
Mean for push	4.64	−0.04	0.03	3.42	−0.06	−1.26
Mean for push and pull	2.53	0.16	0.23	1.44	−1.34	0.56

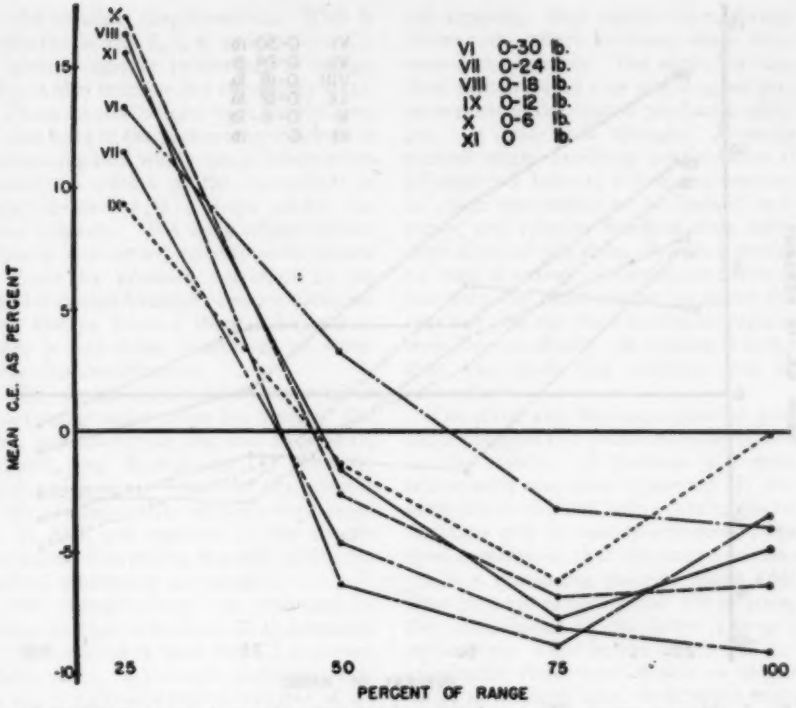

FIG. 4. Mean percentage CE's as a function of percentage of pressure range, maximum displacement constant at 30° (pull)

bined is significant beyond the .01 level of confidence. It should also be noted that for pull, for five of these conditions, the CE changes from positive to negative between the 25% and 50% points of the range. In the conditions where pressure range was constant this change occurred between the 50% and 75% points.

Variable error as a function of pressure.—The *SD*'s of the mean percentage CE for the conditions where displacement range was maintained at its maximum of 0–30°, and pressure range varied from 0 to 0–30 lb., are given in Table 4. No significant changes in the variability around the mean CE of a condition occurred as pressure range was varied. Indeed, the variabilities are strikingly close. As with the conditions where pressure was kept constant, the variability declined as the displacement within a range grew larger.

DISCUSSION

The data, when all the scores made with a particular condition are combined, show that (*a*) with pressure range constant, mean percentage CE and relative variability decreased as displacement range was increased, and (*b*) with displacement range constant, mean percentage CE and relative variability were not significantly affected by changes in pressure range. Furthermore, the differences found between conditions were mainly the contribution of the differences

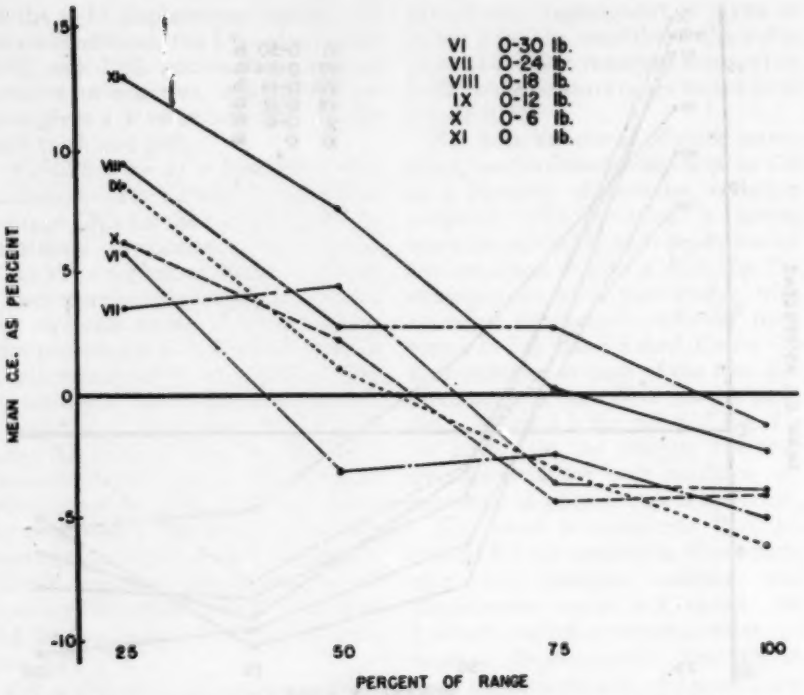

FIG. 5. Mean percentage CE's as a function of percentage of pressure range, maximum displacement constant at 30° (push)

TABLE 4

SD OF PERCENTAGE CE AS A FUNCTION OF PRESSURE RANGE, EXTENT RANGE CONSTANT AT 0-30°

Displacement in Degrees	Pressure Range					
	0 Lb.	0–6 Lb.	0–12 Lb.	0–18 Lb.	0–24 Lb.	0–30 Lb.
Pull						
7.5	28.35	21.31	24.65	28.57	25.12	31.01
15	18.89	13.13	19.93	15.30	17.13	23.66
22.5	14.63	16.83	13.48	13.42	13.91	12.66
30	13.40	14.07	9.01	10.61	7.87	9.34
Pull (combined)	21.79	19.34	18.66	20.87	18.99	22.15
Push						
7.5	23.39	30.54	19.96	25.60	21.45	20.65
15	15.63	15.94	13.14	16.52	15.53	15.70
22.5	13.06	11.60	12.15	12.44	12.90	12.80
30	12.98	10.29	8.07	7.56	9.17	11.89
Push (combined)	17.87	19.43	15.07	17.33	15.88	16.20
Push and pull (combined)	19.98	19.38	16.96	19.28	17.55	19.49

at the smallest displacements. This is indicated in Fig. 2, 3, 4, and 5 where CE is plotted against percentage of range. This is also true for the variability data.

These results indicate that, for the kind of task used in the present experiment—a positioning task where visual information about the effects of the movement is delayed—extent is a more useful cue than pressure. The data cannot reveal whether any contribution to performance is made by pressure variation at the shorter ranges because pressure variation was always present there. An experiment is now being performed to determine this contribution.

The results obtained on accuracy of positioning movements by Spragg, Devoe, and Davidson (9), and by Brown, Knauft, and Rosenbaum (1) generally agree with those found in the present study; namely, that relative error tends to be high and positive at the shorter distances, decreasing sharply as the extent of movement is increased.

The "range effect," as described by Ellson and his co-workers (2, 3), appeared in these results in both the CE and variability data. A distance underestimated in one condition where it was one of the longer distances tended to be overestimated in another condition where it was one of the shorter distances. For example, Fig. 2 shows that a setting of 3° is underestimated in the condition where it was the longest distance (I) and overestimated in the condition where it was the shortest, and 12° was the longest (III). The variability was also smaller in the first case than in the second. The experimental data support the view that these differences are not due solely to pressure differences, if, indeed, these had any influence at all.

The findings of the present experiment conflict with those of Gibbs and Baker (5) and Gibbs (4), who found a pressure control to be superior to a free moving or distance control. Two explanations for this discrepancy arise from the differences in the experimental conditions. The first is that *S*s in the above-named studies were provided with continuous visual feedback, so that, while making their adjustments, they could immediately observe the effect achieved when they moved the control. The second is that their task involved rate tracking, so that movement of the control produced velocity, not positional changes. Pressure control might facilitate performance if adjustments have to follow one another in quick succession or be braked suddenly, and velocity tracking does make these kinds of demands. Tracking guided by cues of extent of movement involves not only the time needed to move the arm but also the time needed to recover from inertial effects. At present, it seems that the conflicting findings can be reconciled.

The data and the experimental procedure suggest two possibilities to account for the results: (*a*) distance is a more informative cue than pressure; (*b*) distance is remembered better than pressure. Distance may be a more informative cue than pressure in that distance variation yields a measuring scale so much finer than pressure variation that the information contributed by the latter is largely redundant. That is, the number of discriminable differences within a usable range of distance may be so much more than the number within a usable range of pressure, that variation in the latter gives *S* no data that he does not already possess in the other form.

Furthermore, distance variation involves cues from several sources: degree of muscular contraction; the pattern of extensor and flexor innervation; and the geometrical relationships of the limb, which, in the task used here, involve information from three joints—shoulder, elbow, and wrist. Pressure variation involves mostly variation in the degree of muscular contraction, or intensity.

In adjusting the control, *S* had to predict what a particular application of force and/or displacement would do to the position of the spot—a prediction based on information derived from his previous settings on that condition. If pressure cues are not remembered as well as distance cues then the former would serve no useful function in the task used in this experiment. Should this be so,

it would help to explain the discrepancy between the findings of this study, where learning and memory played a key role—and those of Gibbs (4) and Gibbs and Baker (5), where the contribution of memory was much less. It should also be noted that the number of sources to which memory for distance might be tied is very likely greater than the number available to memory for pressure.

Summary

The purpose of the present experiment was to determine the accuracy of positioning responses as pressure and distance information was varied, using a task where visual feedback was delayed. Eleven different force-displacement conditions were used to provide the variation in information. The *S*'s task was to compensate for the displacement of a spot of light from the center of an oscilloscope screen by moving a stick control the proper distance and in the proper direction. Eleven *S*s were used, each serving under all 11 conditions.

The results showed that relative error and variability decrease with distance moved but that pressure variation has no apparent effect on the measures. Further, the effect of variation in distance was found to be greatest at the smallest displacements within a condition. It was concluded that, in positioning responses made in the absence of visual feedback, displacement is the more crucial dimension of the force-displacement relationship of a control.

References

1. Brown, J. S., Knaupt, E. B., & Rosenbaum, G. The accuracy of positioning reactions as a function of their direction and extent. *Amer. J. Psychol.*, 1948, **61**, 167–182.
2. Ellson, D. G., & Wheeler, L. The range effect. Dayton, O.: Wright-Patterson Air Force Base, 1951. (*Tech. Rep.* 5813.)
3. Ellson, D. G., & Coppock, H. Further analysis of the psychological range effect. Dayton, O.: Wright-Patterson Air Force Base, 1951. (*Tech. Rep.* 6012).
4. Gibbs, C. B. The continuous regulation of skilled response by kinesthetic feed back. Medical Research Council, A. P. U. Report No. 190/53.
5. Gibbs, C. B., & Baker, J. C. Free-moving versus fixed control levers in a manual tracking task. In A. Tustin (Ed.), *Automatic and manual control.* London: Butterworth, 1951, 467–472.
6. Hick, W. E. The precision of incremental muscular forces. Report No. 23, Applied Psychology Unit, Cambridge University, 1945.
7. Jenkins, W. O. The discrimination and reproduction of motor adjustments with various types of aircraft controls. *Amer. J. Psychol.*, 1947, **60**, 397–406.
8. Orlansky, J. Psychological aspects of stick and rudder controls in aircraft. *Aeronaut. Engng Rev.*, 1949, **8**, 1–10.
9. Spragg, S. D. S., Devoe, D. B., & Davidson, A. L. Studies in the accuracy of movement: II. The bisection and duplication of linear extents at various angles in the horizontal plane, and at inclinations above the horizontal. Port Washington, L. I., N. Y.: U. S. Navy, Special Devices Center, 1949. (*Tech. Rep.* SDC 241-6-1.)
10. Weiss, B. The role of proprioceptive feedback in positioning responses. Unpublished doctor's dissertation, Univ. of Rochester, 1953.

(Received for early publication September 25, 1953)